(1)

AD-A239 843

Technical Report 927

Integrated Knowledge Elicitation and Representation Framework

John Leddo, Marvin S. Cohen, Michael F. O'Connor,
Terry A. Bresnick, and Freeman F. Marvin
Decision Sciences Consortium, Inc.

April 1991

United States Army Research Institute
for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief cf Staff for Personnel

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract for the Department of the Army

Decision Sciences Consortium, Inc.

Technical review by

Michael J. Barnes
Merryanna L. Swartz

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-POX, 5001 Eisenhower Ave., Alexandria, Virginia 22333-5600.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Technical Report 927

Integrated Knowledge Elicitation and Representation Framework

John Leddo, Marvin S. Cohen, Michael F. O'Connor,
Terry A. Bresnick, and Freeman F. Marvin
Decision Sciences Consortium, Inc.

Field Unit at Fort Huachuca, Arizona
Julie A. Hopson, Chief

Systems Research Laboratory
Robin L. Keesee, Director

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

April 1991

Army Project Number
2Q162785A790

Human Performance Effectiveness
and Simulation

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS --
2a. SECURITY CLASSIFICATION AUTHORITY -- 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE --	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) 90-3	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Technical Report 927
6a. NAME OF PERFORMING ORGANIZATION Decision Science Consortium, Inc. 6b. OFFICE SYMBOL (If applicable) --	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute Fort Huachuca Field Unit
6c. ADDRESS (City, State, and ZIP Code) 1895 Preston White Drive Reston, VA 22091	7b. ADDRESS (City, State, and ZIP Code) ATTN: PERI-SA Fort Huachuca, AZ 85613-7000
8a. NAME OF FUNDING / SPONSORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences 8b. OFFICE SYMBOL (If applicable) PERI-S	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-86-C-0383
8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
62785A	790	1306	C02

11. TITLE (Include Security Classification)
Integrated Knowledge Elicitation and Representation Framework

12. PERSONAL AUTHOR(S) Leddo, John; Cohen, Marvin S.; O'Connor, Michael F.; Bresnick, Terry A.; Marvin, Freeman F.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 86/10 TO 89/12	1991, April	

16. SUPPLEMENTARY NOTATION
Contracting Officer's Representative, George Lawton

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Knowledge elicitation Expertise Expert systems Battle management

19. ABSTRACT (Continue on reverse if necessary and identify by block number)
This report summarizes the results of a project designed to develop methods to elicit and represent expert knowledge. Previous work in knowledge elicitation/engineering has been largely applications-driven, with expert knowledge force-fit into predefined knowledge frameworks. The goals of this project were to gain insight into how experts structure knowledge, to develop and test a set of integrated knowledge-elicitation techniques driven by this understanding of expert knowledge representation and use, and to gain insight into the nature of expertise and how it develops. Knowledge-elicitation sessions were conducted with some fifty experts in order-of-battle analysis, situation assessment, target development, and collection management at Forts Bragg, Carson, Lewis, Hood, and Huachuca. Results indicate that experts use a diverse range of knowledge structures depending on their functional area, level of experience, unit, and other variables. These structures are used integratively rather than singly. As a result, we propose an Integrated Knowledge Structure (INKS) framework for expert knowledge representation. We found that using a diverse set of

(Continued)

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL George W. Lawton	22b. TELEPHONE (Include Area Code) (703) 274-5565 22c. OFFICE SYMBOL PERI-ICC

DD Form 1473, JUN 86   Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

ARI Technical Report 927

19. ABSTRACT (Continued)

knowledge-elicitation techniques integratively elicits knowledge more completely than any single technique (such as think-aloud problem solving). We also found that the most experienced subjects reason primarily by causal analysis driven by their goals or objectives. Comparison of experts across different skills levels suggests several stages may occur in the development of expertise. These findings have implications for (1) the development of artificial intelligence, expert, and decision support systems that more closely model expert knowledge, (2) training people to become experts, and (3) evaluating the effectiveness of training by comparing student knowledge to expert knowledge.

FOREWORD

Expertise is a precious commodity. Experts have been shown to outperform novices on many problem-solving tasks. Capturing the knowledge of experts benefits the Army in many application areas: development of expert systems that model expert problem solving, thereby enabling the system to aid or automate problem-solving functions; training, where the goal is to teach students to reason and solve problems as experts do; and personnel selection, where the goal is to select people with appropriate expertise and skill levels for particular tasks.

Because expert knowledge is rich and diverse, a comprehensive set of elicitation techniques is needed to elicit and integrate different types of expert knowledge. In addition, a knowledge-representation framework is needed that is rich enough to capture diverse and rich expert knowledge. This report outlines a set of complementary and integrative knowledge-elicitation techniques, each of which focuses on different aspects of expert knowledge. It outlines an Integrated Knowledge Structure (INKS) framework for representing expert knowledge.

The report focuses on these techniques and the research that was conducted. Appendix A contains a "how to" guide for selecting and implementing the techniques and evaluating the elicited knowledge. Appendix B contains probes used by the different elicitation techniques. Appendix C contains coding sheets that can be used by the elicitor for the elicited knowledge.

This report presents a comprehensive set of knowledge-elicitation techniques. With practice, an elicitor will become proficient in their use, and will be able to apply them to selected applications such as expert system development and training.

EDGAR M. JOHNSON
Technical Director

ACKNOWLEDGMENTS

We would like to thank the project's Contracting Officer's Representative, Dr. George Lawton, for his helpful advice and guidance throughout this project. Dr. Lawton deserves recognition more as a collaborator than a project monitor, and we are grateful for his support. We are also grateful to Dr. Julie Hopson and Dr. David Burnstein for their helpful feedback on the project.

We appreciate the guidance of Dr. Rex Brown. A special thanks is reserved for Ms. Julie Doherty for her help with data organization, scenario generation, and for being there when we needed her.

We are grateful to Brigadier General James Hunt (U.S. Army, Retired) and Major General Edmund Thompson (U.S. Army, Retired) for their help in scenario generation and for acting as co-elicitors.

Finally, our thanks to the many active duty Army personnel who offered their valuable time as subject matter experts. Their contributions produced the informative elicitation sessions on which this work is fundamentally based.

INTEGRATED KNOWLEDGE ELICITATION AND REPRESENTATION FRAMEWORK

EXECUTIVE SUMMARY

Requirement:

Developing models of how experts represent and use knowledge can serve many valuable applications. Given the fast paced, time-stressed battlefield environment, expert and decision support systems will play an important role in future warfare. Understanding how experts solve problems and make decisions will enable the development of systems that show superior performance across a wider range of applications and that are more compatible with their users than systems developed without this understanding.

Further, the Army invests considerable resources training its personnel. Acquiring expertise is a lengthy process and few succeed. As a result, military experts are few in number and their tenure is short. If training methods to help more people acquire expertise and acquire it more quickly were developed, it would be of great benefit to the Army. We argue that understanding how experts represent and use knowledge can serve as a blueprint for training methods to enable more people to emulate experts.

The proposed project had two goals: To develop a set of techniques that elicits expert knowledge, and to develop a knowledge-representation framework that captures how experts represent and use knowledge in a psychologically valid way. This report presents the findings of the second phase of a two-phase project that involved extensive field research with military intelligence specialists who were practitioners in the areas of situation assessment, order-of-battle analysis, target development, and collection management.

Procedure:

The first step in our work was to review the literature on knowledge representation and knowledge elicitation. This led us to develop two preliminary knowledge-elicitation frameworks, interpretative and generative. These frameworks were tested in elicitation sessions at Fort Bragg, Fort Carson, and Fort Huachuca. The results of these sessions are presented in the phase I final report (Leddo, et al., 1988).

The results of the phase I studies led to the development of our Integrated Knowledge Structure (INKS) framework for representing expert knowledge. This drove the development of new techniques that were then tested in phase II. The techniques that comprised the year 2 studies were problem solving, Project Ahead (designed to elicit the knowledge experts use to develop scenarios of potential outcomes of actions), Conflict Resolution (designed to elicit how experts reason with uncertainty), and Cognitive Structure Analysis (designed to elicit the structure and content of expert knowledge structures). These techniques were tested individually in knowledge-elicitation sessions at Fort Bragg and Fort Lewis and then integrately during two separate trips to Fort Hood.

In all, some 30 sessions were conducted during phase II. Most sessions lasted 3 or 4 hours, some a full day, and two were 3-day sessions (these tested the full integrated methodology). The elicited knowledge was analyzed for indications of complementary and complete techniques, and for insights into how experts represent and use knowledge.

We conclude with recommendations and examples as to how our INKS and knowledge-elicitation frameworks can be used in areas of artificial intelligence (AI) and expert systems and also in training and training evaluation. We present some insights into what makes someone an expert and how expertise is acquired.

Findings:

In contrast to most characterizations of expert knowledge in the literature, experts use multiple and integrated knowledge schemes rather than single schemes such as frames or production rules. However, experts may emphasize different aspects of their knowledge, depending upon their domain and their level of experience.

For example, order-of-battle analysis is largely template-driven, while situation assessment is largely plan-driven. The least experienced and skilled subjects seem to rely most heavily on template-based reasoning; the most experienced and skilled subjects seem to reason using goals and causal mental models.

The set of knowledge-elicitation techniques used in this project works well together. Collectively, the techniques capture a diverse range of knowledge skills with little unintended duplication of elicited knowledge (suggesting the techniques work together efficiently) or contradiction. The techniques are fluid in that they allow for easy transition among themselves.

Our experiences with the knowledge-elicitation framework suggest that rather than develop a fixed set of elicitation procedures, it is more useful to develop a set of guidelines for using particular procedures, shifting procedures, and assessing the knowledge that has been elicited and where to emphasize future elicitations. We feel this will better suit the needs of future elicitors, whose goals may not all be the same.

Utilization of Findings:

We see our findings being used in three areas--knowledge-representation applications, knowledge-elicitation applications, and understanding how expertise is acquired. These applications serve the areas of expert, decision support, and AI system development, training, and training evaluation.

We believe using our INKS framework will lead to the development of more powerful AI and expert systems that will be more compatible with their human users. We review one AI system developed at Decision Science Consortium (DSC) that uses the INKS framework to support understanding, explanation, planning, and learning.

The INKS framework can also serve as the expert knowledge model that is the basis for training programs. Such programs would be designed to teach students to emulate experts. We review one intelligent tutoring system, developed by DSC, that uses INKS in such a fashion.

We believe that using our integrated knowledge-elicitation methodology, which is aimed at producing psychologically valid models of expert knowledge, can serve as the basis for building expert knowledge models and designing expert systems. Current expert systems typically define the system architecture first, then force-fit expert knowledge into it. Our approach would be to elicit the expert knowledge and then design the architecture. The result should be a more powerful, more user-compatible system.

The elicitation methodology could also be used in training evaluation. Here, the elicitation techniques could be used to elicit students' underlying reasoning strategies to see how they differ from experts'. This can be used to develop remedial instruction tailored to the specific knowledge requirements of individual students and that point students toward high levels of expertise. We discuss how the Intelligent Tutoring System, referenced above, actually does this.

Finally, our findings can be used to gain insight into what makes an expert and how expertise is acquired. The former is useful both to identify experts in a given domain and to serve as a model for training and as the basis for developing training methodologies.

CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

INTEGRATED KNOWLEDGE ELICITATION AND REPRESENTATION FRAMEWORK

Introduction

Experts, needless to say, are good at solving problems in their domain of expertise. Knowledge engineers seek to fashion formal representations of the knowledge that underlies that ability; such representations are being applied in a growing range of fields (medicine, geology, military intelligence), and for a growing number of purposes (computerized aids, the design of training systems, the development of new decision-making procedures). Knowledge engineers have, from the start, used a pragmatically motivated mix of techniques to "extract" this knowledge: examining written sources, observing experts in actual and hypothetical problems, open-ended probing ("How do you do your job?"), and more specific questioning. To a surprising degree, however, they have relied on experts themselves to formulate if-then statements, or rules, that purportedly characterize their performance.

Despite its importance, knowledge elicitation remains an ill-understood and eclectic art that demands an enormous amount of time from both the elicitor and the expert. The elicitation process, and its attendant difficulties, raise fundamental issues about the data structures and reasoning strategies associated with diverse types of expertise and about the manner in which such knowledge is tapped by different elicitation methods. These issues have not yet been dealt with in a satisfactory way, either in the artificial intelligence community or by cognitive psychologists.

The present research underscores two factors that, we believe, must be acknowledged and dealt with if more efficient elicitation techniques are to be developed.

Knowledge Elicitation as a Constructive Process. The object is not so much to capture knowledge that pre-exists in the expert, as to build, or design, such knowledge in collaboration with the expert. The concept of design emphasizes choices and tradeoffs, and these are forced upon the knowledge engineer in numerous ways:

> Experts, we will argue, have command over a multiplicity of representations and problem-solving strategies. These must be sorted out, their proper roles identified, and apparent inconsistencies resolved.
>
> Some forms of knowledge do not permit easy verbal articulation by the expert. The knowledge engineer must fill in gaps in the data, identify possible

lapses or errors, and extrapolate from the observed set of problems to others that may occur in the domain.

Each elicitation method which the elicitor chooses to work with will introduce the possibility of error or distortion in a different way. The knowledge engineer must select methods, and evaluate their results, in the light of their respective strengths and weaknesses.

Expert judgment may appear to diverge from normatively justified principles for reasoning (cf., Kahneman, Slovic, & Tversky, 1982). The knowledge engineer must decide what is at fault: the experts, the normative principles, or the elicitation methods.

Multiple experts in the same domain may disagree in their methods, in their conclusions, or in both. The knowledge engineer must determine when and how competing views should be reconciled, and when the possibility of alternative approaches, and even multiple possible conclusions, should be allowed for.

The knowledge must be translated into computer implementation (or other presentation) formats that are limited in their ability to capture the flexibility and variety of human concepts and reasoning methods.

Clearly, the metaphor of knowledge elicitation as "mining" for "nuggets" of knowledge (e.g., Hayes-Roth, et al., 1983) is misleading. There may be no static, unconnected "nuggets" waiting to be found. Elicitation is a process of <u>definition</u> and <u>redefinition</u> in which the knowledge engineer models patterns in the observed data (interpolating, extrapolating, and smoothing out) while striking a balance among competing representational constraints (consistency within an expert, between experts, with normative principles, and with an implementation format). The elicitor's queries often are not intended to trigger a canned response; in many instances they present a "problem" to be solved, through selective attention, retrieval, inference, and redescription within formal or informal constraints, by the expert; similarly, the expert's answers often present a "problem" (of interpretation, follow-up, and formalization) to be solved by the elicitor. In such a process, both the expert's and the knowledge engineer's picture of the knowledge may shift as multiple viewpoints are explored.

<u>Expert Knowledge Embodied in Structures</u>. The notion that knowledge is not composed of independent pieces (or "nuggets") was implicit in the above discussion. Constraints imposed by

els, for example, apply only when elements of n be combined to imply other elements of knowledge. ily by reference to such potential combinations that ise to speak of gaps in knowledge to be filled or cies to be resolved. In the constructive approach to , which we urge, the knowledge engineer has one eye on l model and the other on the individual elements. ive nature of elicitation consists in adjusting each to h the other. A pattern discerned among the individual suggests a possible model structure; the model structure, ction with the individual elements, implies certain elements, and the validity of the model can be checked by g the expert's acceptance of those elements (or, possibly, sion to revise one or more of the original elements). on also occurs when multiple models, and their tions, are compared with one another--whether from the same ferent experts.

first blush, this emphasis on structuring and modeling may appear to add to the elicitor's burden compared to proach that seeks knowledge one piece at a time. In fact, opposite is true. Knowledge structures enable experts to de large quantities of information economically by exploiting ndancies, or regularities, in their problem domain. uctures permit them to derive or generate information that is encoded explicitly from information that is. Experts not only e "more knowledge" than novices; they have better theories--d it is those theories that, at least in part, enable them to ive more knowledge. By the same token, knowledge engineers can, nd should, utilize such structures to increase the efficiency of nowledge elicitation by obtaining large quantities of implicit knowledge for every element of explicit knowledge that is elicited.

Normative structures such as those developed by decision analysts (e.g., Bayesian inference, decision trees, multiattribute utility theory) have some of the desired properties, i.e., once certain assessments have been made, others can be derived. But such structures tend to be too problem-specific to serve plausibly as representations of expert knowledge, and a typical complaint in practice is the large number of assessments that are required even for relatively simple problems (e.g., Schum, 1980). To significantly improve the efficiency of knowledge elicitation, more general structures are required, which, when manipulated and combined, generate representations tailored to a specific situation.

In the cognitive science literature, a variety of highly general knowledge structures have been proposed with a descriptive, rather than prescriptive, motivation (e.g., "scripts" by Schank & Abelson, 1977; Memory Organizational Packages (MOPs)

by Schank, 1982). These are offered, typically, as general explanations of broad classes of experimental findings and observed behavior (e.g., generation of expectations in familiar settings); but there has been no systematic account of how they are used in reasoning (i.e., how do they generate implications?). In addition, no guidance has been offered for empirically confirming the presence of such a structure in a particular case, rigorously identifying its components, or distinguishing it from other types of structures. Without addressing these issues, elicitation dialogues in which experts are asked to fill in "slots" in frames remain quite unconvincing.

Later in this paper we will describe a method, called Cognitive Structure Analysis, for systematically identifying structures in expert knowledge, rigorously confirming their validity (by reference to the implications they generate), and fleshing them out. In accordance with our constructive view of elicitation, we do not claim that such structures exactly capture what experts know. They represent idealizations and approximations mutually arrived at and agreed upon by the elicitor and the expert. Nevertheless, they constitute plausible explanations of expertise, and provide a far more effective and efficient approach to elicitation.

Background

Why is knowledge elicitation hard? Among cognitive scientists who have considered that question, perhaps the prevailing viewpoint is that expertise consists, at least in large part, in the automatization of performance. Anderson (1982), for example, has proposed that the development of cognitive skill involves a shift from explicit declarative knowledge of domain-specific facts to implicit procedural knowledge of those same facts.[1] Others (e.g., Ericsson & Simon, 1980) have argued that people do not have direct awareness of causal processes in their behavior, but only the actual contents of working memory. Putting these two claims together, we have a syllogism:

(a) Expertise is embodied in process, not in content,

(b) People have no direct awareness of causal processes in

[1]Thus, a sophisticated novice, when he learns that "object a is F," might explicitly think about a generalization of the form "all F's are G's;" a general-purpose (non-domain-specific) rule then causes the conclusion, "a is G," to be drawn. By contrast, after he becomes an expert, the presence of "a is F" in working memory causes "a is G" to be activated directly, without any consideration of the explicit generalization ("all F's are G's"). The latter has been replaced by a domain-specific rule or procedure (if you know "a is F," then conclude "a is G"), of which there is no explicit representation in working memory.

their behavior, and

(c) Therefore, experts have no direct awareness of the knowledge that constitutes their expertise.

Emphasis on automaticity leads to an avoidance of elicitation techniques in which experts are asked to articulate general knowledge underlying their performance; Ericsson and Simon, for example, favor think-aloud problem solving, in which experts are asked merely to verbalize the immediate contents of working memory. Think-aloud instructions are said to avoid errors introduced by the redirection of the expert's attention, by the recall of past events, or by the influence of stereotypical causal beliefs on the expert's efforts to generalize.

The impact of these arguments on knowledge engineering practice has not matched its acceptance by theorists. While think-aloud problem solving is often used, it is seldom employed in the pure form envisaged by Ericsson and Simon; experts are still being asked to state or validate general rules. We would argue that there are good theoretical and practical reasons for this resistance; in fact, that both premises in the syllogism are, at the very least, misleading. Considering why will enable us to revisit the two principle themes of this paper. Discussion of premise (b) will lead us to an appreciation of the sense in which knowledge elicitation is a <u>constructive</u> process. Consideration of premise (a) will then highlight the importance of knowledge <u>structures</u> in elicitation.

Premise (b) imposes a crisp distinction between the immediately presented, absolutely certain contents of consciousness and what can only be inferred. Yet think-aloud problem solving is itself not a direct window onto the expert's problem-solving competence. What the expert reports may reflect a variety of distortions: (1) verbalization may compete for cognitive resources with the initial problem-solving task and may alter or degrade problem-solving performance (Berry & Broadbent, 1984); conversely, the problem-solving task may reduce the completeness with which memory contents can be verbalized; (2) the choice of what to verbalize need not accurately reflect the factors that are governing problem-solving-task performance; and (3) theories and prior expectations may influence the way an expert describes what he is thinking (especially when his thoughts involve non-verbal imagery).

Implicit in the argument based on premise (b) is the notion that inference can and should be avoided in knowledge elicitation. Think-aloud problem solving does not eliminate inference; it shifts the principle burden of constructing a general knowledge representation from the expert to the elicitor. To build a production-system representation, the knowledge engineer must make a series of conjectural and mutually interdependent decisions about segmenting the protocol, filling in apparent gaps,

specifying the possible knowledge states of the expert, identifying the processes which operate those knowledge states, and identifying the conditions for their applications (see Waterman & Newell, 1971).[2] Production system models that are based on think-aloud data (e.g., Newell & Simon, 1972) invariably draw upon the elicitor's own pre-existing knowledge of the task domain (e.g., chess, cryptarithmetic) in order to narrow down the vast range of representational choices left open by the data. It is unlikely that an elicitor could even get off the ground in the absence of such prior domain knowledge of this sort.

Ericsson and Simon's arguments against relying on directive, general querying of the expert can thus be turned on their head, and used equally against reliance on general models (inferred from noisy and incomplete data) by the knowledge engineer. We would argue that this is an instance of a more general rule: to the degree that we ease the burden on the expert to articulate and shape his own knowledge into the desired form, we increase the burden on the knowledge engineer, to fill in gaps, extrapolate beyond the given data, and reconcile apparent inconsistencies.[3] Note in addition that these latter activities make sense only by reference to an elicitor-selected mode of representation (e.g., a Production System) and its associated constraints. Once the crucial role of top-down judgment in knowledge elicitation is appreciated, it becomes pointless for the knowledge engineer to deny himself the benefit of the expert's own general judgments regarding the domain unless one is prepared to claim (as premise (a) apparently does) that the expert has nothing to offer. We

---

[2]Suppose the think-aloud protocol contains the comments: "a is F...a is G." Even in this simple case, the expert cannot conclude there is a rule "If anything is F, then it is G." There may be other, unverbalized features of a that were necessary (so the correct rule might be "If anything is F and H and I, then it is G."). Conversely, the most economical representation may involve a more general antecedent ("If anything is L, then it is G" where the F's are a subset of the L's). Finally, of course, the expert may intend no connection between F and G at all (G being derived from a set of unverbalized preconditions) or if there is a link between F and G, it may be indirect ("All F's are Q's; all Q's are R's; all R's are G's."). Enormous quantities of problem-solving data would be required to sort out all these possibilities, for a reasonably rich domain, in a "bottom-up" fashion.

[3]Identical tradeoffs arise in psychophysics, where measurement techniques differ in what is demanded from subjects (more precisely, the scale of properties of data that are exploited) and, conversely, in the degree of latitude permitted in modeling the data. The same tradeoff--between the use of think-aloud protocol analysis and numerical judgments to model choice--is implicit in Einhorn, Kleinmuntz, & Kleinmuntz (1979).

will argue shortly for a notion of expertise involving multiple representations: that expertise grows not only toward an increasing repertoire of specific prepackaged responses, but also toward an increasing ability to reason abstractly and flexibly about the domain. Expertise may evolve separately (but interdependently) along these two paths: increased "tuning" (discrimination and generalization) of automatic processes (Anderson, 1982) as well as simpler and more powerful theories.

The more valid conclusion, then, is to give up the quest for a single perfect technique, as well as the assumption that there is a static store of knowledge in the expert's head waiting to be "extracted." A combination of elicitation methods, e.g., think-aloud problem solving plus more directive and general questioning of the expert (in which the expert is asked, in effect, to help "model" his own knowledge) may yield improvements in both efficiency and validity by providing checks and balances against the limitations and biases of each individual method.

We turn now to a more direct consideration of premise (a) and, in particular, evidence suggesting that it is not the whole story. This evidence suggests that expertise involves the development of qualitatively different knowledge structures and reasoning processes, and not simply an increasing quantity of if-then rules, or increasing automatization of existing rules, as suggested by premise (a). Automaticity has its primary effect on the speed of performance, by collapsing large numbers of operations into single routines. Improvements in performance are by-products, resulting from the ability to generate a larger number of possible solutions in a shorter period of time. Moreover, there is a self-limiting character to the improvements created by automaticity: the increasingly specialized response packages they produce have a correspondingly decreased range of useful applications (Newell & Rosenbloom, 1981).

There is evidence that the superiority of experts over novices goes beyond the rapidity with which operations are performed. In the first place, experts utilize different entities and properties to organize their understanding of the same facts. In physics, Chi et al. (1981) found that novices judge the similarity of problems on the basis of "surface structure," i.e., superficial features such as type of apparatus, while experts judge similarity by reference to basic principles of physics (e.g., conservation of energy) and generic solution techniques associated with such principles. Similar differences between experts and novices in algebra are reported by Shoenfeld and Herrmann (1982) and in computer programming by Weiser and Shertz (1983). The clusters in recall found by Chase and Simon (1973) for chess masters suggested that meaningful patterns within the chessboard were classified at an abstract level, e.g., as "attack" configuration or "defense" configuration. McKeithen, et al., (1981) and Adelson (1984) obtained similar results for recall clustering by expert and

novice computer programmers. Finally, the entities mentioned in expert as compared with novice think-aloud problem solving protocols tend to be technical rather than familiar, and to be closely tied to fundamental laws (Larkin, 1981; Greeno, 1983).

Most importantly, the entities and properties utilized by experts appear to be associated with knowledge structures and reasoning processes that are both more general and more flexible than those of novices. Larkin (1981) identifies a number of aspects in which expert inferences are more robust than those of novices: e.g., (a) properties of objects, as perceived by experts, are local rather than defined in terms of expected effects on other objects, thus permitting use of knowledge about these objects across a wider range of contexts; (b) inference by experts involves atemporal constraints among properties as opposed to single-direction temporal simulation, thus permitting the same property to serve as an input in one problem context and as an output in another; and (c) the same problem can be solved in more than one way, utilizing different sets of input properties.

Although it is formally possible to represent this kind of inferential ability in a rule-based system, a simpler and to us more persuasive structure is provided by representations based on schemata or frames (cf., Larkin, et al., 1980; Novack & Araya, 1981; Greeno, 1983). If-then rules specify the use to which the knowledge is to be put; thus, each potential use requires a separate representation. As a result, although rules are claimed to be "modular" pieces of knowledge, in fact such a system widely disperses knowledge about the same object or concept, making changes difficult. Schemata collect knowledge about a particular object or concept in one place, and the same representation can be utilized in a large variety of different ways (Winograd, 1975).[4]

---

[4]For example, one type of schematic representation, the script (Schank & Abelson, 1977), represents a sequence of events commonly associated with some activity or goal. An extremely simplified example from the tactical miliary domain might be the following <u>attack</u> script:

ATTACK

Move Artillery Forward	Mass Troops	Artillery Attack	Troop Movement Across Forward Line

In situation understanding, if <u>any</u> one of the specified events is observed, the others can be inferred (with some degree of uncertainty): e.g., after a massing of enemy troops is observed, we may infer (a) that artillery have been already moved forward, and (b) that an artillery attack and forward troop movements will occur in the future. Similarly, if troop massing and an artillery

Rule-based and schema-based representations reflect contrasting criteria of efficiency. Procedural representations maximize the speed of a specific response to a specific stimulus (since there is, in effect, a direct S-R connection). Schematic representations, on the other hand, provide significantly more compact representations of a large amount of knowledge: a single schema (together with a small number of general-purpose procedures for interpreting it) takes the place of numerous specialized rules. Schemata thus have the same advantages as simple theories in science: they account for a wider range of phenomena with fewer postulates. By the same token, learning by acquisition of schemata should be far more efficient than learning by acquisition of rules.[5]

The latter point has vital implications for knowledge elicitation. Perhaps the single most significant source of inefficiency in knowledge elicitation is the commitment to representational "modularity," i.e., ostensibly independent chunks of knowledge typically in the form of rules (Ericsson, Simon, & Anderson question only the traditional reliance on an expert's formulation of these rules). This commitment leaves the elicitor with the task of "extracting" tens and perhaps hundreds of thousands of separate pieces of information from the expert in a reasonably complex problem domain. And it implies that the expert himself has gained his knowledge (from direct exposure to the domain) in the same piecemeal fashion. We think that in practice neither the expert nor the elicitor acquires knowledge in this way. Both of them, whether consciously or unconsciously, learn and exploit structures that remain hidden in a rule-based representation.

---

attack have been observed, we infer (a) that forward movement of artillery has already taken place, and (b) that troop movements are imminent. By contrast, in a production rule representation, separate rules would be required for each possible combination of inputs.

[5]A schema can in fact be represented compactly as a small cluster of if-then rules relating observable stimuli to some theoretical entity or property (e.g., if artillery is moved forward, and attack will occur;...if attack is to occur, then: first, artillery will be moved forward; second, troops will be massed; third,...). But meta-rules will be required for interpreting these rules and extracting all their implications. In effect, then, the first order set of rules is a declarative and not a truly procedural representation, i.e., it is data, not process (and it is this feature that makes possible the representation by a small number of rules).

If this is right, then knowledge elicitation is hard not because there are so many rules to acquire, or even because they are "automatic," hence, inaccessible to verbalization by the expert. Elicitation may be difficult in large part because elicitors have tried to capture knowledge piecemeal, and have failed to appreciate and go after the structures that enable experts to organize and utilize large quantities of information.

From this point of view, our expectations regarding the expert's ability to articulate his expertise are somewhat changed, and no longer a matter of black (process) and white (data). We think the equation of "data" (or explicit representation in working memory) with "verbally describable" is misleading. An activated knowledge structure is not automatically verbalizable; the ability to describe it will depend on the degree to which words are incorporated within or linked to the schema, or the degree to which the schema matches general-purposes verbal procedures, or skills, of the expert. A mix of techniques--ranging from think-aloud problem solving to more specific verbal questioning--may be required to fully elicit and characterize such structures.

An Integrated Knowledge Elicitation Methodology

Much of the previous work in knowledge elicitation and knowledge engineering has been applications driven. Most knowledge elicitation efforts are used to establish knowledge bases for expert systems. Here, the goal of the elicitor is to generate a knowledge base which is procedural (i.e., can be replicated by a computer) and functional (produces output similar to that of an expert). Typically, the representation of such knowledge is driven by the programming environment, e.g., whether the system is rule-based or frame-based, with production rules being the typical framework of choice.

Unfortunately, the expert system approach to knowledge elicitation largely ignores one of the more promising applications of knowledge elicitation technology, namely gaining insight into how experts organize and use knowledge. While it is true that experts often exhibit reasoning behaviors which appear "rule-based" (cf., Newell & Simon, 1972), our research shows that experts use diverse modes of reasoning suggesting that different types of knowledge and reasoning skills are present (Leddo et al., 1988).

We would contend that utilizing knowledge elicitation techniques to uncover the diverse modes of reasoning and knowledge representation that experts use would increase the power and flexibility of the resulting expert systems which embody this knowledge. We believe that by predefining the representation

framework and force-fitting elicited knowledge into it, knowledge elicitors and expert system designers have limited the power and usefulness of their systems.

Our approach to knowledge elicitation, then, can be viewed as more "exploratory." Rather than predefining a product (e.g., an expert system design) and a representation format, we use our elicitation techniques to "discover" the different ways in which experts represent and use knowledge. This approach places a special demand on the elicitor. S/he must have elicitation techniques which are flexible enough to handle diverse and even unanticipated modes of knowledge use, yet powerful enough to pursue these different modes of reasoning in detail. As a result, the elicitation process itself must become fluid and dynamic. This is contrasted with the standard "think aloud problem solving protocol format" where the expert is given a problem, is asked to solve it while articulating his thoughts, and the resulting protocol is then converted into a series of production rules. We argue that this format offers very little in the way of active probing or exploration of the expert's knowledge.

Such a standard approach may be adequate if the elicitor merely wishes to replicate the expert's output in like situations. However, if the elicitor wishes to gain insight into the reasoning skills and knowledge structures that produce successful problem solving so that s/he can develop a more powerful, richer knowledge model that generalizes across problem sets, the elicitor must be able to go beyond the superficial behavior observed in the "think aloud" technique. We argue that no single knowledge elicitation technique is powerful enough is capture the different types of knowledge and skills that experts have.

Our approach to knowledge elicitation has been to develop a set of complementary techniques that aim to get at different modes of reasoning and knowledge representation. These techniques were driven by reviewing the empirical literature on expert reasoning and knowledge representation. We then attempted to develop techniques that individually capture major knowledge representation and reasoning skills and collectively complement each other. While we do not claim that our techniques capture all of expert knowledge, we believe we have taken a significant theoretical step in knowledge elicitation technology by: (1) viewing the expert, not the application, as the driving force in the knowledge elicitation process; (2) viewing the expert as using diverse methods of representing and using knowledge; and (3) viewing the elicitation process as an exploration (much the same way an archaeologist puts together pieces of information he "digs up," using his knowledge of the domain and the information s/he's gathered so far to guide his future exploration).

The theoretical starting point for our integrated knowledge elicitation methodology is our INKS ((Integrated Knowledge Structure); Leddo, Cardie, & Abelson, 1987; Leddo & Cohen, 1987; Leddo, Sak, & Laskey, 1989; Leddo et al., 1988) framework for knowledge representation and use. We summarize the INKS framework here.

In the following paragraphs we refer to the different types of knowledge that experts have. These include goal and planning knowledge, knowledge of situation-specific procedures, knowledge of data patterns, content knowledge of objects, causal models of how events are produced and high-level integrative knowledge.

In the cognitive psychology and artificial intelligence (AI) literatures, a number of representation schemes have been proposed as models of human knowledge representation. These schemes tend to address specific types of knowledge. For example, scripts/MOPs (Schank, 1982; Schank & Abelson, 1977) are used to represent goal and planning knowledge that occurs in fairly routinized environments. Script-like schemata can also be used to integrate bodies of knowledge into a larger framework.

In the literature, scripts and MOPs have been described as generic planning structures, associated with specific contexts, and are organized hierarchically around key goals. The heart of these schemata are scenes which describe discrete action sequences (often associated with specific people and sub-contexts) designed to meet specific subgoals of the schema. For example, if we view the situation development process as a script or a MOP, where the goal is to determine enemy course of action, then this might give rise to a subgoal of locating key enemy elements. In turn, this might organize a scene "visit the collection manager" whose purpose is to provide guidance on collection elements. This scene might organize the actions involved with the collection manager, e.g., going to the shop, conveying intent, providing priority intelligence requirements (PIRs), finding out what assets are available, etc.

Interestingly, the collection manager may have his own collection management schema that could contain the scene "meet with the G-2." This scene may describe the same interactions described above, but from the collection manager's perception (of course, there should be substantial overlap).

In addition to the scenes, scripts and MOPs have additional information, some of which is structural, some of which is background. In terms of structure, the goal relationships determine sequencing and location in the hierarchy of the actions in the schema. Background information includes the conditions under which the schema is fired ("entry conditions"), the people involved ("roles"), the expected outcome, and the tools or props used.

Knowledge about data-patterns and how objects are organized together (e.g., how units make up a motorized rifle division) can be represented by frames (cf., Anderson, 1980; Minsky, 1975). Frames are very much like scripts in that they are expectancy-driven organizers of knowledge. In our conceptualization, scripts focus more on goal and plan-related knowledge while frames organize collections of objects. In addition, scripts and MOPs seem to organize temporal knowledge while frames seem to organize spatial knowledge. Frames can be viewed as templates or patterns of stereotypic arrays of objects. For example, how a tank division aligns itself on offense is an example of a frame.

As far as we can tell, there is no fixed set of slots for a frame. The slots seem to be more a function of what is being represented. However, there are some commonalties across frames. For example, frames typically have slots for the parts or objects that make up the frame. In addition, any spatial arrangement of these objects would also be contained in the frame. Frames may also contain categorical relationships, e.g., battalions are part of regiments, tank divisions are ("isa") combat unit. Frames may also contain information regarding their purpose, e.g, tank units are exploitation forces.

Frames can also be distinguished from semantic nets (cf., Quillian, 1966) which tend to organize information about individual concepts and relationships between them rather than collections of objects. For example, a tank division may best be represented by a frame since it is a collection of tanks, artillery, etc., while a single tank may best be represented by a semantic net which describes the features a tank (e.g., its speed, its physical make-up, etc., as well as examples of different types of tanks).

Semantic nets have two basic components, nodes and links. Nodes are specific concepts such as tank and gun while links describe particular relationships between the nodes (e.g., tank has-part gun). Typically, the semantic literature has focused on two types of links, those relating to categorical relationships and those relating to properties or features of objects. The primary categorical relationship is "isa" where an object is an instance of a more general category (e.g., Abrams "isa" tank). The properties links can be more varied depending upon the different properties of an object (e.g., speed, range, lethality of weapons).

Knowledge about situation specific procedures can be represented by production rules (cf., Newell & Simon, 1972). Production rules are expressed in the form "IF [antecedent], THEN [consequent]", where antecedents are situational conditions which determine when procedures are to be executed and consequences are the procedures executed under those conditions. Production rules

are useful both in carrying out actions/plans (e.g., "If this is the situation, then do these actions") and in also generating inferences (e.g., "If the following data are observed, then infer that this is the situation"). Production rules can be distinguished from scripts in that scripts organize entire goal-driven plans, while production rules organize specific actions. Scripts can be viewed as collections of production rules much the same way that frames can be viewed as collections of semantic nets.

Finally, causal and analogical reasoning can be captured by mental models (cf., DeKleer & Brown, 1981; Johnson-Laird, 1983; Leddo, Cardie, & Abelson, 1987; Leddo & Cohen, 1989). In our framework (Leddo, Cardie, & Abelson, 1987; Leddo & Cohen, 1989), mental models are as viewed mechanisms that underlie events. Mental models include information regarding how forces and objects interact to produce outcomes.

In the literature, there has been no established template or fixed set of slots for mental models. This is so primarily because mental models have been used to represent real world or conceptual processes (e.g., how electricity works). As such, the form of the mental model was patterned after the phenomenon it was modeling. Indeed, the mental model was structured so as to be isomorphic to the phenomenon in question.

We have adapted the mental model to focus primarily on causal understandings of processes or events. As a result, we have developed a set of slots that we feel are necessary in order to understand a causal process. First, the mental model contains knowledge of the goal or purpose of the model (for example, if the mental model concerns "momentum" the expert knows that the goal of the model is to achieve an objective or victory). Next, the mental model contains the objects involved in the process or event. These objects can be physical objects (e.g., weapons), people, etc. The mental model contains the forces involved. Forces act on objects to produce outcomes. Such forces can be physical (e.g., gravity) or motivational (e.g., desire). Next, the model contains knowledge about how the forces and objects interact. This is the heart of the mental model and describes how the process works (in this respect, much of the previous work in mental models may fall into this slot). Finally, the mental model contains an outcomes slot that describes what happens when the forces and objects interact. These outcomes often relate back to the mental model's goals.

Mental models can be used in an explanatory mode (e.g., this outcome happened because these forces acted upon these objects). Mental models can also be used in prediction or plan generation (e.g., if we exert this force on this object, this outcome will occur).

We have discussed five different representation frameworks (scripts, object frames, semantic nets, production rules, and mental models) for representing expert knowledge. Leddo et al. (1988) has shown that experts possess knowledge that can be represented by each of these frameworks. It is important, then, to integrate these separate frameworks into a larger, more general framework of knowledge representation. In our INKS framework, scripts serve as the general organizer of knowledge, linking plans and goals together. Production rules give situation specific procedures to be executed given conditions that arise during the execution of a plan. Frames organize collections of objects that are utilized in the execution of plans while semantic nets organize features for the individual objects within a frame. Mental models provide the causal rationale for why procedures are executed and how they are instrumental in achieving objectives.

For example, a representation for the process of situation development, using our integrated representation framework, might start with a general situation development script. This script would depict the general goals of situation development, e.g., determine enemy's intent, capabilities, etc. and plans for achieving those goals. These plans would include general procedures such as discussing named areas of interest and information requirements with the collection management and dissemination section, interacting with the all source production section and the intelligence analysts of adjacent, higher and lower units, etc. Within these plans would be inference rules (expressed in production rule format) for how to process information (e.g., if an enemy division moves its artillery forward, then this is an indicator that the enemy is preparing to attack).

The representation would also include frames that are referenced by the script and/or production rules, e.g., if a reference is made to "Soviet tank division," a frame of that type of division is activated. Similarly, semantic nets are linked to the concepts contained in the frames (e.g., tanks) as well as to concepts referred to elsewhere but not contained within a particular frame. Finally, mental models may be linked to production rules or general plans and procedures to give underlying rationales. For example, a mental model generated by one of our subjects for why artillery moving forward indicates attack is that the general flow of the battle will be the attacker advancing and the defender retreating. Hence, the artillery is moved forward to be able to range a retreating defending force.

Given this general framework for knowledge representation and use, we now illustrate how this would drive the development and selection of knowledge elicitation techniques. Our goal is to develop a set of complementary techniques that collectively will capture as complete a set of knowledge skills and representations as possible.

Given that the heart of our INKS framework is goal and planning knowledge of the kind characterized by scripts, the central technique of our integrated knowledge elicitation methodology should capture that type of knowledge. Based on this reasoning, the think aloud problem solving technique (see Leddo, Cohen, Bresnick, & Marvin, 1989 for a detailed discussion) forms the heart of our methodology. We have taken the standard think aloud technique and created variations of this in order to capture different aspects of the problem solving process. For example, full-scale context driven problem solving is used to capture general problem solving procedures. Data-driven problem solving is used to elicit situation-specific knowledge associated with specific problem solving and inferential procedures.

During the course of problem solving, experts often project future states of the world (environment) that they might encounter so that they can plan for these situations. Such projections often involve the causal mental models discussed above. These causal models enable the expert to predict future events (e.g., the outcome of a battle) given the present situation (e.g., deployment of forces along the terrain) and anticipated causal factors (e.g., the forces that will be involved in a battle). A technique that we have developed to elicit these "projections" is called "Project Ahead." Briefly, the Project Ahead technique has experts imagine various scenarios that might evolve from a particular situation. The expert is led through a variety of "branches" in the scenario (i.e., different ways in which the situation could evolve depending on how the environment or the enemy reacts), including being asked why the events would unfold the way they would. The Project Ahead technique is discussed in some detail in Brown (1989) and as well as in Appendix A.

While the expert is assessing these future situations and in general conducting his problem solving analysis, s/he must often consider data that is coming in and how it relates to the current situation and the expert's objectives. Here, the expert is asked to draw inferences and conclusions from potentially incomplete or unreliable information. Hence, the expert must know how to handle uncertainty in his analysis. We have developed a technique called "Conflict Resolution" that is designed to elicit these types of knowledge. Briefly, the expert is presented with pieces of information and asked to draw inferences or conclusions from them. The expert is asked under what conditions the inferences would be true and what conditions would invalidate the inferences. The expert is also challenged to examine his thinking by being instructed to assume in one case that the premises to his argument are true, but the conclusions false, and in another case that the conclusions to his arguments are true but the premises are irrelevant. The expert is then asked to explain this conflict. We hypothesize that this latter challenge sheds important insights into the different ways an expert can approach a problem, since

the elicitor attempts to "invalidate" the expert's standard problem solving approach and force him to generate new approaches. Hence, the expert comes up with new solutions.

Both the Project Ahead and Conflict Resolution techniques are designed to help flesh out some of the problem solving mechanisms which experts use. We believe that this results in a more fine grained, richer analysis than the standard problem solving protocol technique. However, using this trio of techniques (Problem Solving, Project Ahead, Conflict Resolution) still focuses on problem solving behavior (albeit at a richer level) without addressing the underlying structures that give rise to these behaviors. While it is arguable (and perhaps correctly so) that one can never "truly" elicit "the representation" that the expert has, we believe that a good stab can be made at some of the knowledge relationships that the expert has that facilitate these types of behaviors.

For example, Galambos (1986) presents research that suggests that script-like knowledge is organized hierarchically rather than sequentially. His conclusion was based on the finding that speed of access of knowledge in a script did not depend on its temporal sequence in the script, but rather on its importance. For example, in a restaurant context, "eating the meal," which is more important but occurs later in the script is accessed quicker that "asking for the menu," which is less important but occurs earlier.

Based on this type of reasoning, we developed a fourth technique called "Cognitive Structure Analysis" (Leddo & Cohen, 1989). Cognitive Structure Analysis (CSA) is a question and answer (interview) technique designed to infer the structural aspects of knowledge representations by looking at their logical implications (as in the Galambos study). CSA has two major steps: identifying the structural aspects of the knowledge being elicited and filling in the content of the structures.

To summarize, the four techniques comprising our integrated knowledge elicitation methodology are Problem Solving, Project Ahead, Conflict Resolution, and Cognitive Structure Analysis. As discussed earlier, Problem Solving is the heart of the methodology, its goal being to elicit the primary goal and planning knowledge that experts have. Problem Solving is supported by the Project Ahead and Conflict Resolution techniques. Project Ahead elicits the knowledge the expert uses to project out the sequence of events s/he is likely to encounter, while Conflict Resolution looks at the way the expert reasons with the data s/he is given during problem solving. CSA supports these techniques by eliciting the structural aspects of the knowledge the expert uses and fleshes out details not covered by the other techniques. These individual techniques and guidelines for their use are discussed in greater detail below, followed by discussions of how

they may be integrated and how they compare in efficiency. Finally, results are described regarding the use of different knowledge structures in different functional areas.

Elicitation Using Problem Solving

One of the key factors that distinguishes experts from novices is their ability to solve problems. In fact, researchers such as John Anderson (1982) argue that the development of expertise centers on proceduralizing more declarative knowledge about how to solve problems. As a result, the most widely used knowledge elicitation techniques center around having experts solve realistic problems. Typically, experts are asked to articulate their thoughts as they work through the problems. These "protocols" are then used to construct representations of the expert's problem solving knowledge.

We view problem solving-based knowledge elicitation as a set of techniques that can be used depending upon constraints such as the elicitor's goals, the characteristics of the domain, and the kinds of experts available to serve as subject matter experts. We discuss them briefly here and present in Appendix A more detailed guidelines for deciding when to select which technique variations and how to conduct knowledge elicitation using the different techniques.

We begin our discussion of problem solving techniques (we use the term "problem solving techniques" to mean problem solving-based knowledge elicitation techniques rather than techniques experts use to solve problems), by first describing some of the key types of knowledge the knowledge elicitor might be focusing on. This in turn would drive the characteristics of an appropriate knowledge elicitation technique, which in turn would drive how that technique is constructed and used. It is our goal to present not only a set of knowledge elicitation techniques that we have already developed but also a conceptual model of the knowledge elicitation process, which a knowledge elicitor could then use to construct his own techniques, if necessary.

In general, problem solving involves selecting and applying plans and procedures to achieve specific goals. We view the problem solving process as having two major components: an assessment of the problem situation to identify the appropriate goals, causal elements, features, etc., that both define the problem and suggest an appropriate problem solving strategy, and; apply the problem solving strategy to arrive at a solution. The first component is largely an inferential task and may involve the use of low-level content knowledge and more general pattern oriented knowledge (cf., Chase & Simon, 1973). The second component is more procedurally oriented and involves linking goals

and plans to specific situations. From a global Army perspective, these two components are analogous to the G-2 and G-3 functions, respectively (the G-2 does situation assessment/inference and the G-3 does planning and plan implementation), although we argue that both of these functions occur (although differently) at all problem solving levels.

We argue that the first component, problem/situation assessment, is largely a "bottom-up" process, i.e., the expert must link low-level content knowledge about a specific problem situation to a higher level pattern or problem solving strategy. Our Phase I results (Leddo et al., 1988) suggests that this type of reasoning was prevalent among order of battle (OB) experts whose job it was to put together pieces of data to form a definition of the situation.

We argue that the second component, plan selection and implementation, is largely a "top-down" process, i.e., the expert must apply a general, often abstract plan to a specific situation. Hence, more generic procedures and rules of thumb must be concertized and instantiated in a specific situation.

Elsewhere (Leddo, Cardie, & Abelson, 1987; Leddo & Cohen, 1987, 1989; Leddo, Sak, & Laskey, 1989; Leddo et al., 1988), we describe our INKS framework for representing the different high- and low-level knowledge and its integration that enable successful problem solving. We review the major highlights here.

In our INKS framework, knowledge is organized largely around goals and plans. Experts tend to have a set of goals that characterize their domain (e.g., situation developers want to know enemy intention, composition, strength, disposition, etc.) and a set of plans for achieving them. Experts know what procedures to execute in order to implement the plans successfully, including how to differentiate the procedures across different situations. In addition, experts have rules and guidelines for discriminating when and how to select and use plans across different situations. These decisions are often driven by the features of the current situation and how they causally relate to the objectives the expert is trying to achieve. Hence, experts have a great deal of domain-specific content knowledge and causal models which relate the content and plans to the relevant objectives. This content knowledge is often packaged into patterns or "frames" which represent frequently occurring situations. As a result, experts often develop sets of inference rules for recognizing when these patterns occur.

Given the analysis of the problem solving process and key knowledge components, we can discuss variations in the problem solving technique that might be more appropriate for eliciting different portions of the complete problem solving process. As we have discussed, problem solving involves two broad tasks: problem

assessment and plan implementation, the former defined as developing the problem "picture," the latter as applying the problem solving strategy.

We argue that most of the work done in problem solving has focused on the latter, i.e., plan implementation. Here, experts are often given a well-defined problem and asked to produce a solution. The military context is different from most domains in that problems are typically ill-defined and a large part of the problem solving process is trying to assess what the problem (situation) characteristics are.

We propose two broad categories of problem solving techniques: context-driven and data-driven. Context-driven problem solving is similar to the standard problem solving approaches used in common knowledge elicitation practices (cf., Ericsson & Simon, 1984). Here, subjects are presented with a complete, relatively well-defined problem and asked to produce a solution. We argue that this type of problem solving approach is useful in eliciting problem solving strategies and plans. This knowledge is apt to be largely procedural in nature (this is consistent with typical findings in the knowledge elicitation literature where knowledge engineers present experts with context-driven problems and build production rule/procedural representations of the experts' protocols).

Data-driven problem solving (Leddo et al., 1988) explores the first component of problem solving, namely, developing a picture of the problem. In doing so, experts must process the features or elements of the problem, assess what factors are relevant, and infer how the different features relate to each other and the problem goals. This type of knowledge is likely to be more inferential (based on inference rules) in nature and involve things like pattern recognition and causal diagnosis. The data-driven problem solving involves presenting the expert with a stated goal (e.g., develop an assessment of the enemy's most likely course of action), but minimal context. The expert is then fed individual pieces of information and is asked to process them and integrate them to form a picture of the problem. More detailed procedures for carrying out context-driven and data-driven problems are presented in Appendix A.

Another dimension which characterizes problem solving knowledge is its level of abstraction. For example, planning structures tend to be general in nature and applicable across a wide range of contexts. For example, the Army's Mission Enemy Terrain Troop-Time Available (METT-T) planning strategy is meant to be applicable to all tactical situations. As such, there is little in the way of situation-specific procedures to execute--rather it is a general way to approach tactical problems. The expert must fill in the specific procedures using his content knowledge of the specific problem (e.g., enemy being fought, terrain, etc.).

One way to restrict the knowledge that an expert will use to solve a problem is to present the expert with highly-constrained and specific problems to solve. Such problems should require a specific, limited task be performed in a well-defined problem area (as opposed to a well-defined problem). We term such problems "mini-problems" since they focus on a well-defined subset of more general problem solving knowledge. Mini-problems force the expert to focus on a specific, small-scale problem which requires the use of instantiated procedures (i.e., the general planning knowledge has to be instantiated into a set of specific, executable procedures). Mini-problems, therefore, focus on eliciting specific procedural knowledge as opposed to more abstract, generic planning knowledge.

Hence, our second major problem solving technique dimension (which is orthogonal to the dimension of context-driven versus data-driven problem solving) is whether a full-scale or mini problem is used. Full-scale problems involve presenting the expert with a broad problem which captures many aspects of the job he performs (e.g., present a situation development (SD) expert with a full-scale scenario and ask him to develop an estimate of the situation). Mini-problems involve presenting the expert with a small subset of the general problem and observing an in-depth solution to the problem (e.g., the same expert is asked to compute a force ratio analysis of the friendly versus enemy forces).

We argue that full-scale problems will induce the expert to solve the problem at a more general level, using relatively generic problem solving strategies. (Indeed, our experts typically glossed over the details of the problem.) Mini-problems will induce the expert to focus on problem details and use more situation specific problem solving strategies. Appendix A presents guidelines for conducting knowledge elicitation using mini-problems and full-scale problems.

A final dimension that is relevant, particularly in the military context, is the number of problem solvers. Most knowledge elicitation has focused on single expert problem solvers or using multiple experts to develop a consensus (often called "decision conferencing"). However, often problems are solved using inputs from multiple sources and these inputs need to be coordinated and integrated. This is especially true of military decision making which is distributed in nature.

Group decision making settings require a different type of knowledge on the part of experts. The expert must not only know how to perform his own job, but he must have a sense of how his job fits in with the organizational perspective. He must know how to integrate what he is doing with the work of others, he must understand the requirements of his fellow group members, he must know how to use the group's resources (both people and materials)

to meet his own needs. Indeed, many of our intelligence experts (particularly, the collection managers) cited "knowing how to work the system" as a major aspect of their job.

It is difficult to conduct a complete elicitation of a group decision making process using single expert problem solving. While the single expert can provide a good picture of the analytical tools and processes he uses, the organizational/group dynamics are best elicited using group problem solving techniques. Hence, single versus group problem solving is the final dimension that differentiates our problem solving techniques. Guidelines for conducting knowledge elicitation using group versus single problem solving is presented in Appendix A.

Methods

Subjects. Subjects were 13 officers, ranging in rank from warrant officer to lieutenant colonel, recruited from Fort Lewis and Fort Hood. These subjects represented functional areas of order of battle analysis (OB), situation development, target development (TD), and collection management (CM). Of the 13 subjects, 4 were recruited from Fort Lewis (I Corps and 9th ID) and represented one each of the four functional areas. The remaining 9 subjects recruited from Fort Hood (III Corps, 2d AD, 1st Cav) and were comprised of 4 collection management specialists, 3 target development specialists and 3 all-source analysts. Two explanations are needed. First, 10 "specialists" have been generated from 9 subjects. This is because one subject, who was a specialist in both all-source analysis and target development was run for two independent sessions, one for each specialty. Second, based on research prior to the Fort Hood trip, it was learned that many of the situation assessment and order of battle functions were performed by the same people in the all-source production section. Hence, to accommodate the sample size, available subjects, and knowledge elicitation conditions to be used at Fort Hood, OB and SD were collapsed into a single function: all-source analysis. The task of all-source analysis was defined as determining the location of the enemy main attack, a primary concern of both OB and SD.

Scenario Used. The elicitation session centered around a standard Fulda Gap scenario. While the combat units at Fort Lewis have missions in the Pacific rather than the European theater, all subjects at Fort Lewis had experience with the European problem, either from previous assignments or from classroom training. The primary mission of the units at Fort Hood is the theater in question.

The scenario used in the sessions depicted a U.S. division (the 52d MID) in a defensive posture against a Soviet Combined Arms Army (the 8th CAA), with a separate Tank Army (7th TA) as a

potential follow-on force. The 8th CAA was comprised of three divisions, two relatively attrited Motorized Rifle Divisions (MRD's) in contact with the 52d MID and at nearly intact 9th Tank Division in reserve. The scenario had two variations: a "standard" problem version (similar to problems typically presented in training exercises) and a "novel" problem. These are discussed in turn.

The "standard" scenario began with a lull in the fighting. The enemy had been conducting limited probing attacks and the reserve forces (9th Tank Division and 7th TA) had been advancing toward the Forward Edge of the Battle Area (FEBA) (the boundary between friendly and enemy forces). The 9th Tank Division had settled into an assembly area approximately 15 kilometers (km) from the FEBA and the 7th TA had been continually advancing and its lead elements were approximately 50 - 60 km from the FEBA.

Subjects were provided a variety of sources of information. In particular, there were 1:50,000 and 1:250,000 maps depicting division and Corps, respectively, areas of interest. For each of the maps, there were overlays depicting friendly and enemy OB as of the morning of the current day's fighting. Since the subject was solving the problem from the division's perspective, most of the information provided related to things the division would be interested and to the 1.50,000 map. The subject was also given overlays of the terrain analysis of the division's sector and intelligence preparation of the battlefield (IPB) products in the form of overlays depicting named areas of interest (NAIs), target areas of interest (TAIs), and an event template.

In addition to the graphic information, subjects had notebooks containing detailed information on enemy OB, a database of relevant events in the war, relevant Corps and division mission statements, weather and terrain information, and a summary of enemy capabilities and vulnerabilities. Finally, there was a message stream which presented processed intelligence information in the form of hourly intelligence summaries (ITSUMS).

The "novel" problem was designed to force subjects out of a standard problem solving mode. It represented three days later in the fighting described in the standard scenario. Here, the enemy forces had penetrated in the north and were attacking the friendly rear area, while friendly forces had penetrated in the south and were attacking the enemy rear area. The intelligence information was largely incomplete so that it was unclear which enemy units had made the penetration (e.g., the 71st MRD of the 8th CAA, the 9th Tank Division or elements of the 7th TA). In addition, the message streams presented with the novel scenario suggested that the enemy was using a new weapon, possibly laser-based.

Procedure. Several variations of the problem solving procedure were used. The variations included single vs two person problem solving and full-scale vs mini-problems. All subjects received context-driven problem solving. Subjects at Fort Lewis received both the standard and novel problems, while subjects at Fort Hood received standard problems only. Unfortunately, there were insufficient subjects to run a complete factorial design and vary task (OB, SD, TD, and CM) with problem type (standard vs novel) and problem solving technique variation (single vs group problem solving, full-scale vs mini-problem, context- vs data-driven problem solving). A full design would have required four tasks by two problem types by eight problem solving technique variations or 64 conditions requiring subjects. Given the limitations on subject availability, it was decided to run key conditions and sacrifice statistical significance on any given condition.

At Fort Lewis, four subjects were run, one from each of the four functional areas. Each subject was run individually (single problem solving) and received the context-driven, full-scale standard problem. Subjects in the role of SD and OB then received the context-driven, full-scale novel problem (for other subjects, there was insufficient time to work through both problems). The SD subject also performed a mini-problem (compute combat power ratios of friendly versus enemy forces). Sessions lasted from three to four hours depending upon the availability of the subject's time.

At Fort Hood, nine subjects were run, four collection management experts, two target development experts, two all-source analyst experts, and one subject who doubled as an all-source analyst expert and target development expert (this subject had over twenty years of experience and was the most expert of all the subjects). These subjects were run in six session, four of which were in group problem solving conditions. Of the four groups, one involved the pairing of an all-source analyst with a collection manager, and the remaining three groups involved the pairing of two experts in the same functional areas. Of these three groups, one group represented each of the functional areas of all-source analysis, target development, and collection management. Of the experts run individually, there was one each representing the functional areas of target development, collection management, and all-source analysis. The one subject who served as subject matter expert for two domains was run as part of the two all-source analyst group problem solving and as the single target developer.

Two of the six sessions had an embedded mini-problem. The group representing collection management was given a mini-problem of collecting against the 7th TA (the Front's follow-on force in the scenario). The single subject representing target development expertise received a mini-problem of prioritizing artillery targets.

Sessions at Fort Hood ranged from 2 1/2 to 4 hours, depending upon the availability of the subject matter experts.

All subjects were given a briefing on the background to the problem. This briefing was provided by an co-elicitor who had a military background. The co-elicitor was available to answer technical questions the subjects had about the scenario. Subjects were then given a specific task. All-source analysts (including SD and OB experts) were asked to develop an assessment of where the enemy main attack is most likely to come, collection managers were told to develop a collection plan (in the case where a collection manager and an all-source analyst worked together as a team, the collection manager developed a collection plan to support the all-source analyst's assessment of where the main attack would come), and target developers were asked to select targets.

Subjects then worked at their own pace. They were encouraged to think aloud and articulate their thoughts. Occasionally, the elicitor interrupted to ask clarifying questions or to urge the subjects to speak up.

All sessions were tape recorded and an on-line transcriber typed the interactions and verbalizations. The transcriber recorded for gist, rather than verbatim. In addition, the elicitor took notes of the interactions and verbalizations.

Results

Data Used. The data used here represents an amalgamation of the three sources of data: elicitor's notes, on-line transcription, tape recording. Essentially, the elicitor went back over the on-line transcriptions and revised them based on the session notes and tape recording. These revisions focused on filling in gaps and correcting errors.

Once the transcripts were revised, they served as the primary source for extracting elicited knowledge.

Full-Scale Problems (Using Single Experts). Full-scale problems are perhaps the best way to elicit a complete pass at problem solving knowledge. Our results suggest that full-scale problems do not elicit the same depth as mini-problems (see section below) but are excellent at capturing the breadth of the problem solving process. Further, with effective probes, such elicited knowledge can suggest ways in which the expert himself represents the knowledge. We illustrate an annotated knowledge summary of one of our outstanding target development experts. This knowledge is expressed in our INKS framework.

"Target Development (MOP-oriented representation since it emphasizes plans).

"Goal: Develop 10-15 good targets for FSE per 24 hour period.

"Step 1 - Revise target list.

"Typical Corps targets: railroads, bridges, road junctions, Army CP (suggests semantic 'isa' relationships).

"Typical division targets: artillery, air defense, division and regimental CP's (suggests semantic 'isa' relationships).

"If enemy river crossing needs to be done, then target engineer assets (suggests production rule).

"Step 2 - Task Collection Assets for Information on Targets.

"General Rule: Task discipline, not collection system (except ELINT).

"Task ELINT to find counter mortar, counter battery radars for AD.

"Plan is to find AD by locating radars through signals emitted.

"Regimental AD: Locate gun dish and dog ear radars.

"Division AD: Locate landroll and straight flush radars.

"COMINT: Focus on locating tube and rocket artillery (priority MRL, then 152 mm. where main attack will be), AD, regimental and division CPs, engineer assets with emphasis on bridging assets, location of non-committed enemy units. Focus on 2d echelon regiment's where main attack might be and second echelon Army.

"SLAR: Use SLAR along roads that might be used by Army and Front second echelon units to locate them and where they are going. First priority is second echelon Army, second priority is second echelon Front.

"Air Recon: Locate second echelon army.

"MI BN: Focus on maneuver units of second echelon regiments where main attack projected. Focus on arty assets, primarily where main attack projected.

"Step 3: Analysis of incoming information.

"1) Take information regarding area target is in and draw target area.

a) Information comes as an ellipse with coordinate, angle and length of short and long axis (suggests semantic information based on properties).

b) Draw ellipse using this information. Ellipse gives 50% probability of target location.

"2) Use doctrine and terrain to find best place for target. Arty likes cover, so look for woodline or town to be near. Also, arty needs area of about 500-700 meters X 200 meters (exception: in World War II, Soviets pt arty in trees and used first shells to blow open holes in trees).

"Step 4: Make a fire plan based on analysis of incoming information.

"Three modes of targeting: arty, jamming, attack by maneuver.

"General priorities:

"For blue MRL: enemy MRL, enemy arty, enemy command and control.

"For jamming: AD, command and control for regiments and divisions.

"For air: armored targets--first priority tanks, second is BMPs.

"Plans:

1) Catch 2nd echelon of Army in assembly area. Actions: target AD with arty so can fly helos to attack division in assembly area.

2) Target division command and control using maneuver elements.

Rationale: Division command and control has three parts:

Division CP
Tactical CP (located forward)
Alternate CP (stays silent so is hard to locate. Comes up if main CP is destroyed or moving (potential frame due to collection and configuration of CPs).

Army also has forward CP which can take over if other three are destroyed.

Therefore: it is hard to get all four CPs. Hence, targeting them by maneuver is not likely.

3) Target enemy artillery using maneuver elements.

Rationale: Arty has many back up systems (potential semantic net).

If take out enemy counter mortar and counter battery fires, the can use sound range.

To take out arty command and control, need to hit BN fire direction center and each battery--four targets total.

Can degrade arty by taking out a battery, but to cause the arty BN to be down for more than one hour, must take out all four targets.

Therefore, it is not practical to attack by maneuver elements.

4) Hit convoy carrying nuclear munitions (this is a high payoff target).

Actions: Hit with air and arty. Use maneuver unit to follow-up and seize or destroy target.

5) Target Army forward CP (normally operates within friendly division's area of interest).
Actions: Attack with air assault. Then use Lance or maneuver units to seize to get documents.

6) Target deep by shaped penetration.

Actions: Hold flanks steady. Have middle units fall back at least 20 kms. (30-60 km. is possible). Corps units attack along FLOT to cut penetrating units off and get into enemy rear.

"Step 5: Pass Data to FSE (Fire Support Element).

"Give six digit coordinates; eight is overspecified. Also target is spread out enough so that six digits is enough."

The above knowledge summary represents a diverse range of knowledge. The expert alternated between assessing goals, to proposing different plans to accomplish these goals, to evaluating the feasibility of the plans. During the discussion, the expert often referred to content knowledge (e.g., configurations of enemy units) which impacted the feasibility of plans.

In essence, the expert's knowledge seemed centered around the objective of selecting targets for the fire support element (FSE). This is consistent with our INKS representation scheme. The

expert's knowledge was organized into broad steps for achieving his goal (analogous to scenes in a MOP, Schank, 1982). However, within each broad step, there were often multiple plans that could be used at that juncture. This represents a departure of the classical way in which MOPs are viewed. Classical MOPs typically organize single generic plans to be applied toward a variety of situations. However, Schank's MOP theory was meant to describe the knowledge of the average person. We believe that one of the characteristics that define experts is that their knowledge becomes much more highly differentiated than novices. Hence, while novices may have limited plans for achieving their objectives they have a wide range of plans available which are applicable across different nuances in the problem situations. The knowledge summary of the above expert supports this hypothesis.

Another noteworthy aspect of the expert's knowledge was that the plans themselves appeared to have a built-in analysis of the feasibility of their use. This analysis was a knowledge hybrid of content (semantic/frame) knowledge, with causal reasoning relating to planning procedures. For example, the expert's analysis of targeting enemy command and control using maneuver units included his knowledge of the components of the enemy command and control system, procedural knowledge regarding how he might carry out such an attack, and a causal analysis of what would happen under those circumstances. This suggests that experts' plans may be more than simply a list of procedures to be triggered when certain conditions are met (which is how plans are typically viewed in the artificial intelligence (AI) community.

First, plans may be more integrative in the knowledge they contain. This is consistent with our INKS approach to modeling knowledge. Second, plans appear to involve not only procedures with causal reasoning to project plan outcomes, but also a set of evaluative criteria and/or analysis procedures for assessing whether plans meet desired objectives. We have seen this evaluative interplay in the planning process with other experts and believe that it is an important and integral part of the planning process. The evaluative knowledge appears to come into play in cases where there are no canned plans and the expert is left to improvise. As a result, experts must often adjust or "tweak" their plans in order to gain maximal effectiveness. We believe this interaction between procedural, causal, and evaluative knowledge is of great interest and believe it deserves more attention in the research area of expert knowledge.

Mini-Problems. Several mini-problems were presented to subjects. The OB/SD functional area received a mini-problem concerning the generation of force ratios. We use this as an example of the use of mini-problems in knowledge elicitation.

Force Ratio Analysis. Below is a summary of the knowledge elicited on force ratio analysis using the mini-problem.

"Process is done jointly with G-3 and G-2. First, divide area up into divisional boundaries. Draw avenues of approach coming in through FEBA.

"For red side, do count along each avenue of approach for artillery, tanks and APC's. This is done by taking current estimated strengths of units and multiplying by full strength numbers. For example, if current strength of battalion is two-thirds strength and full strength battalion has 30 tanks, then current strength battalion has 2/3 x 30 or 20 tanks. Do this for 1st echelon units along each avenue of approach. Then do same process for friendly forces. Compute separate ratios for tanks, arty and APC along each avenue of approach by dividing number of enemy tanks by friendly tanks along each avenue of approach, number of enemy arty by friendly arty, etc.

"You may tell CDR what type tanks, etc. enemy has, but do not worry what type it is for purposes of ratios--assume all tanks equal, etc. For arty, just count number of tubes, don't look at caliber or range. For APC's count anti-tank (e.g., dragon, stinger and red eye) as APC. Don't count anti-tank as tank [elicitor's note: one subject who was asked to compute combat power ratios counted anti-tank as tanks since anti-tanks kill tanks], since tanks do more than kill tanks--they scare the hell out of infantry.

"Variation in combat power ratio computation: use BN equivalents (takes into account different size BNs in enemy vs. friendly forces, i.e., U.S. units are bigger than Soviet units). Which formula used is matter of preference.

"Evaluating force ratios: 3 to 1 advantage in favor of enemy means defense is tenuous. 2 to 1 advantage in favor of enemy means you are comfortable with defense. Depends on terrain and how defensible--who has dominant terrain (e.g., ridge line or river that channels enemy into one place). G-3 actually makes decision whether he thinks can withstand enemy attack given force ratios. Artillery ratio most important, then tanks, then APC's. If terrain was more defensible then APC's might be higher in importance because infantry can fight in rugged terrain and clear away strong points for tanks. If there is dominant terrain, then arty is more important because can blow out strong points. (Expects Soviets to use their arty to blow out friendly strong points). If terrain is open (and can bypass obstacles and have cross country mobility--which bad weather hampers), then tanks are more important.

"Air combat ratios important, especially if one side had constant air superiority. Air support would be used in helping attack. If one side had superiority, might use air combat power ratio as fourth consideration (in addition to arty, tanks and APC's).

"Integration of terrain analysis with combat power ratios:

"Terrain dictates how much force can send through. Look at how many tanks can put abreast at any given road. The more you can put across, the more fire power you can generate. Assume enemy will use best avenues of approach (i.e., try to generate most firepower) that lead to his objectives.

"Integrating second echelon: Do evaluation by sector (e.g., north or south objective). Assume enemy second echelon will split along avenues of approach in sector. However, compute ratio for sector rather than just for individual avenue of approach. For red side, compute combat power for combination of 1st and 2nd echelon forces in sector (i.e., for each sector use red forces already there and assume that whole second echelon will come there). For blue side, compute combat power for units already in sector plus blue reserve (assuming blue reserve can be committed in either sector). Compute ratio for each sector and separately for artillery, tanks and APCs by dividing Red 1st and 2nd echelon forces by blue on-line and reserve forces.

"Do not count follow-on forces (e.g., 7th TA) until committed in your sector. For this information, go to Corps. Expect 7th TA not to be committed for 24-48 hours because even though lead elements about 50 km away, whole unit is spread out probably close to 100 km. Even though they can move 25 km. hour, when they get close they will have to refuel and get into attack formation before committed to battle. Also, expect blue air support to interdict LOC they're on."

The above elicited knowledge is highly procedural in natural. In fact, it is very formulaic. The procedural nature of the above knowledge may suggest a production rule-type representation. One characteristic of a production rule representation is that it is sufficiently specified so that the procedure could be replicated (either by a computer or a human). To verify this, we gave a naive subject (with no military background) the above knowledge summary and asked him to work on the same mini-problem as the expert from which this knowledge was collected.

The subject was able to run through the expert's procedure and compute force ratios for the given avenues of approach. The subject was allowed to ask questions when he came across a trouble spot. The subject needed the following supplemental knowledge to perform the force ratio analysis.

First, the subject needed to know what were the divisional boundaries. He was able to infer avenues of approach given the terrain analysis overlay. He needed to know which forces were the enemy and which were the friendly. Also, he needed instruction on how to read the unit symbols (e.g., what was the symbol for a battalion, which were the motorized rifle units versus tank units versus anti-tank, etc.). Essentially, this knowledge was background that all military experts implicitly have. Therefore, we did not worry about eliciting it and it is understandable that a naive subject would need this information.

The remaining two knowledge gaps are more substantive in nature. First, the subject needed to know how many tanks and armored personnel carrier (APC) were in each type of unit (we did not have the subject compute artillery force ratios since it would not add anything qualitative to the test of the elicited knowledge) and what were the unit strengths. The expert that originally generated the knowledge also needed this information when presented with the mini-problem. In essence, this is problem-specific information which is not generic to the force ratio formula.

Finally, the subject had a question about where to assign units that did not fall along a specific avenue of approach. This represented an actual gap in the elicited knowledge as this issue was not raised with the expert during the elicitation. For purposes of allowing the subject to complete the analysis, we instructed him to group outlying units with units belonging to the same regiments (i.e., assume the regiment as a whole would fight along a common regimental avenue of approach). Because of this gap in the knowledge summary, we do not consider it appropriate to evaluate whether the subject's answers matched the experts because part of the knowledge the subject used came from the experimenter (i.e., group regimental units together).

However, this evaluation of the mini-problem is highly useful. First, it shows where gaps in knowledge are and in particular, what further knowledge needs to be elicited. Part of this knowledge represents genuine procedural gaps (e.g., what to do with outlying units) and part of this knowledge relates to implicit knowledge taken for granted (but something an expert system would require). Finally, the fact that the subject was able to replicate the expert's process fairly well suggests that mini-problems are useful for eliciting procedural knowledge in depth.

We note one final observation on mini-problems. In the knowledge summary above, we noted a difference in how different experts compute force ratios. The expert described above considered tanks and anti-tanks to be different since tanks can fight infantry, while another expert equated tanks and anti-tanks because both kill tanks. The mini-problem techniques appears valuable at eliciting differences in procedures used by different experts. If probed on these differences, the elicitor can get at the rationales behind these differences.

<u>Group Problem Solving</u>. At Fort Hood, we explored two variations of group problem solving: using two experts performing the same functional areas and using two experts performing complementary functional areas. The most meaningful difference is comparing the knowledge elicited from the group comprised of experts from complementary functions versus comparing the knowledge elicited from the groups comprised of common functions. The group comprised of complementary functions was made up of an all-source analyst (OB technician), while the same function groups were two target developers, two all-source analysts, two collection managers.

It is interesting to compare the all-source/collective management (AS/CM) group with the CM/CM group. The CM/CM group immediately went to the development of a collection plan. This plan represented a generic collection plan which did not appear to relate to the specific problem. For example, the collection management wanted to send long range surveillance unit (LRSUs) out 10-20 kilometers using the aviation brigade (bde) and place one team along each avenue of approach, assign ground surveillance radar (GSRs), communications intelligence (COMINT), and electronic intelligence (ELINT) assets to the brigades and let them control the assets, and fly quickfix up and down the forward line of own troops (FLOT).

On the other hand, the AS/CM group seemed to focus more on the particulars of the situation. First, they developed a set of priority intelligence requirements/information requirements (PIRs/IRs). Then they converted them to indicators (indications of the relevant information) which are then used to generated SIRs ((specific information requirements), i.e., the specific pieces of information that will identify if the indicators are present. An example might be, "Tell me if you see a column of 20 or more tanks moving north on highway 7". This information might serve as an indicator that the enemy was planning to attack in the north.). This then drives the collection plan. It was apparent that in this group, the collection process was being driven by the "user," i.e., the all-source production section. This complementary function group illustrated well the integration of the different functional areas and in particular demonstrated how the "user" of the information drives the information gathering process. (We saw

this same issue in Phase I at Fort Bragg, where the G-2 of the 82d Airborne Division's problem solving centered around providing the G-3 with the information he needed to develop a plan (Leddo et al., 1988).) In essence, this group illustrated how the main goal of the CM function is to provide intelligence analysts with the information they need to put together an assessment of the situation.

In order to gain support for the notion that the mixed function group elicitation process did emphasize the "user" needs and the complementarily of functions, one of the single problem solver transcripts for collection management was compared to the group collection management transcript.

The single problem solver was somewhere in the middle between the two problem solving groups. On the one hand, he deployed his own division collection assets in much the same way as the CM/CM group did. However, when passing collection requests to Corps, he developed sets of indicators based on PIRs and IRs. This may be because when division CM+D passes requests to Corps CM+D, the division CM+D becomes the user of the Corps information (the division CM+D is the dissemination hub for information so it would then be responsible for passing the information that Corps collects onto the all-source production section (ASPS)). Hence, perhaps the "user" focus of the collection process became salient with the request to Corps. Unfortunately, this is speculation based on limited data. However, preliminary data with group problem solving may suggest that when the group is comprised of complementary functions (particularly when the member of the group is a user of the other's product), the "user" or organizational aspects of the functional area are highlighted.

Using multiple experts of the same functional area is useful as well, particularly if the experts are used to working as a team. Such teams demonstrate two types of problem solving skills: consensus building and distributed problem solving. The group made up of two all-source analysts was particularly good at illustrating this.

In terms of distributed problem solving, the two experts divided their efforts initially and later pooled their assessments. The initial work was devoted to information gathering which was done distributively, while the later analysis (i.e., developing an assessment of the location of the enemy main attack) was a joint process. It was interesting to note that during this joint process, the more junior expert deferred to the more senior expert. Occasionally, the junior expert would correct a factual error by the senior expert, but she never contradicted his reasoning or conclusions. We have noticed this one other occasions: a brigadier general deferring to a major general, a sergeant deferring to a captain. Clearly, this is an issue the elicitor must confront when selecting experts for group problem solving and has strong implications on consensus building.

Discussion

We have explored different variations in the problem solving knowledge elicitation paradigm, and in particular, what types of knowledge each variant seems most appropriate to elicit. In general, we hypothesized that context-driven problems would be most appropriate in eliciting general planning knowledge while data-driven problems would be most appropriate in eliciting more low-level inferential and content knowledge (this hypotheses appeared to be confirmed in Leddo et al., 1988--the Phase I study), mini-problems would be useful in eliciting depth of problem solving procedures, and group problem solving would be useful in eliciting the more organizational and interactive aspects of the problem solving process.

Results from the full-scale context driven problem solving suggest that this technique is useful at eliciting a reasonable pass at problem solving breadth. This technique appeared especially useful at eliciting the diverse plans that experts have available to them for problem solving. In particular, this technique was valuable at eliciting the types of reasoning and evaluative processes that underlie plan selection.

Results from the mini-problem suggest that this technique is very useful at fleshing out the details that comprise target plans. The detail achieved by the mini-problem elicitation was sufficient so as to allow replication (with some assistance) of the process by a naive subject. The mini-problem was also particularly useful in pointing out gaps in the elicited knowledge,as evidenced by the questions that the subject had to ask. W feel that this is the most robust test of knowledge completeness, i.e., whether the problem solving process can be replicated. We note that in the full-scale problem, insufficient detail was elicited to allow for replication of knowledge. This may have been due, in part, to the limited amount of time available for the elicitation (we should note that elicitation for expert systems, which require knowledge replicability, often takes months while we were limited to hours). However, it is unclear whether, if we had more time to elicit depth, we would have to use mini-problems to elicit knowledge details or whether the expert would spontaneously narrow his own problem scope to provide detail, thus emulating the mini-problem variation.

Results from the group problem solving suggesting some interesting insights, both in the same function and different function variations. A key difference between the two variations is that different function group problem solving highlighted the user requirements aspects of the experts' jobs and how they drive the process (this is also a benefit compared to single expert problem solving). We believe that a valid model of expert knowledge must take this into account (Martin, Mullin, & Leddo, 1989, make a similar point).

Results from the same function group problem solving emphasized different aspects of group problem solving. Of interest here was the collaborative effort, i.e., how the two experts work together. Our elicitation sessions suggest that experts break up and work separately during information gathering portions of the problem solving process and work together during the analysis portion. In addition, we noted (not surprisingly) that more junior experts defer to more senior experts. Collectively, we feel our results suggest that the different problem solving techniques work complementarily rather than redundantly. Therefore, we feel that the knowledge elicitation process is enhanced through a combination of techniques rather than through use of a single variation. We propose the following sequence (which is repeated in Appendix A which describes how to use the various techniques).

First, the elicitor should begin with group problem solving using experts of complementary functions. The rationale behind this is that the expert's job typically serves an organizational function and this drives what the expert does. Hence, it is important to understand what the expert does from this perspective. The elicitor should begin with a full-scale, context-driven problems. The elicitor should then switch to single expert problem solving. This will enable the elicitor to map out the general domain and the processes the expert uses. The elicitor can then utilize the data-driven technique, again with full-scale problems and single experts. This will flesh out some of the content knowledge that underlies many of the procedures the expert uses. Next, the elicitor may wish to select specific portions of the knowledge base for enhancing its detail. Here, the use of mini-problems (both context- and data-driven, with single experts) may be useful.

We believe that the problem solving knowledge elicitation techniques are some of the richest for capturing expert knowledge. They appear useful in capturing the range of expert knowledge (organization, procedural, causal, etc.) and give insights into knowledge representation which could drive future research. In Appendix A, we discuss how the problem solving techniques can be integrated with other knowledge elicitation techniques to enhance the elicitation process.

Elicitation Using Project Ahead

Methods

Subjects. Subjects were four warrant officers recruited from III Corps, Fort Hood and its subordinate units. One subject represented target development expertise and the other three represented all-source analysis expertise.

Materials. The Project Ahead (PA) technique involves projection of future (or past) scenarios. Hence, the PA technique tends to be situation based. Situations can be described abstractly (e.g., "Tell me how a war in Europe might evolve.") or with a concrete problem (e.g., paired with the problem solving technique). In this study, PA was done both ways, paired with the Fulda Gap problem used in the problem solving technique and with the Cognitive Structure technique (see below).

Procedure. Of the four subjects, two all-source analysts were run as a pair while the remaining two subjects were run individually. Total sessions lasted approximately three to four hours with the PA portion typically lasting about an hour.

As discussed in Appendix A, there are several variations to the Project Ahead technique. PA can be used to project scenarios backward and forward in time. These scenarios can be likely sequences of events, novel/unlikely sequences or the elicitor can choose to put other constraints on the scenario or leave it open-ended for the subject to choose.

We present as an illustration of this technique, the session done with the all-source analyst who was run individually. The subject was asked to generate two likely scenarios that could emerge from the Fulda Gap problem, then asked to generalize likely attack sequences from those scenarios, and finally asked to generate a novel or unlikely sequence of events. This was done to illustrate the flexibility and power of the technique. In addition, the subject was asked to provide a list of indicators he, as an intelligence analyst, would expect to see as these events unfolded.

Results

Data Used. The data used here represent an amalgamation of the three sources of data: elicitor's notes, on-line transcription, tape recording. Essentially, the elicitor went back over the on-line transcriptions and revised them based on the session notes and tape recording. These revisions focused on filling in gaps and correcting errors.

Once the transcripts are revised, they served as the primary source for extracting elicited knowledge.

Analysis. As discussed, the goal of the PA technique is to gain insight into how the expert projects scenarios in specific situations. These scenarios may drive how the expert then processes information and sets collection priorities (see Tolcott et al., 1989).

The text below is taken from the transcript of sessions with the subject matter expert described above. It illustrates some of

the probing sequences in PA. "I" denotes the interviewers queries, while "E" denotes the expert's responses.

> I: "Let's assume that they're (the Soviets) going to attack you, but they're going to do something very novel, but still reasonable...something that you wouldn't expect based on what you know about their doctrine."
>
> E: "The 9th (the enemy second echelon division) withdrawing could be something."
>
> I: "Why would they withdraw?"
>
> E: "Why would they withdraw? Maybe, I, mean, I wouldn't withdraw, first of all, but you want me to come up with something. If the 9th withdrew, maybe the U.S. forces in this sector could be determined to be invincible, and they (the Soviets) just cannot conduct successful operations, or they intend on conducting nuclear operations against us, and they're gonna write off the low percentage units up in the front and they do not want to commit this tank division (the 9th) immediately. They're going to withdraw him which would probably confuse us a great deal, and get him back to a good position where he's protected, fire up the nuclear weapons systems, chemical weapons systems, whatever it is, and then run him forward, right through us."
>
> I: "Let's say they were doing that....what else would you be seeing in conjunction with all this?"
>
> E: "We'd probably see extensive digging in along the front lines of all the troops that are in the present positions, constructing overhead cover, pulling out full NBC gear, probably full-protective masked suits. Taking down their antennas, all their low communications."
>
> I: "What else would you see?"
>
> E: "You'd see guarded convoys moving forward. Security convoys. Probably extensive weather data being taken, meteorological radar would be rather accurate."
>
> I: "Is it going to be in the whole area or just your sector?"

E: "It could be just in our Army's area, maybe they were just selected in one particular area because U.S. forces were determined to be too strong, or else maybe some other NATO allies were not as strong."

I: "Does that make sense? My understanding is that once any one goes nuke, it's fair game for anybody to go nuke."

E: "That could be the case, or it could be selective release for whatever they determine the reason to be. Maybe they'd seen that we saw the 7th tank (army) and the 10th (Combined Arms Army) coming up and that we started to get ready to fire nuclear weapons, and they're getting ready to fire a nuke also in response to our preparation."

I: "What you said could be chemical also."

E: "Yes."

I: "Not necessarily generating nuclear escalation."

E: "Once again, if it was chemical, they could use that right across the front, which would cause us to keep our heads down, and move on through."

Based on the above sequence, it then becomes possible to construct a picture of the scenario that was described. Below is a summary of that event sequence:

EVENT: USE NUCLEAR OR CHEMICAL WEAPONS TO ENABLE BREAKTHROUGH (entry conditions: blue force too strong to break through by conventional means or believe blue force prepare to use nukes).

(Location: May be in Army sector only).

Withdraw follow-on force to safe area (purpose: confuses blue force and keeps follow-on force safe).

Front line troops dig in, construct overhead cover, put on NBC gear, take down antennas, and low communications.

Guarded convoys move forward.

Extensive weather data taken, probably through meteorological radar.

Fire weapons (purpose: weaken blue force and keep blue force heads down).

Run enemy through blue force.

The above sequence captures the events described by the expert. No explicit probes were presented to determine what type of knowledge structure this might represent. However, this sequence would serve as an excellent starting point for such probing. Here, the Cognitive Structure Analysis technique (see below) might be used to determine whether the above sequence may be a script/MOP or other type of structure. Given the sequence of actions and the purposes associated with them, a preliminary assessment is that the sequence is somewhat schematic.

What is interesting about this piece of elicitation is that while the nuclear attack sequence does appear schematic, its application in this context appears to have been created in the session. The subject, when asked to generate something novel, came up with "withdraw" which flies in the face of Soviet doctrine which is offense-oriented. However, when asked to explain this action, the subject seemed to stumble at first and then seemed to come up with a "reason," i.e., they were planning an attack with nuclear weapons. The resulting events appeared doctrinal in nature.

An alternative approach the subject could have taken would be to generate a novel sequence of events or a novel plan involving withdrawal (e.g., induce blue forces to attack and have prepared defense waiting for them or force Blue to waste fuel chasing Red over the battlefield). This may suggest that even in novel situations, subjects have a tendency to look for old plans that they can apply to new situations (or infer them when it's the other side's plan), rather than try to generate new plans in new situations. This is similar to what we found when we used the problem solving technique with a novel problem--subjects still applied old plans to the new problem.

While our findings here do not constitute hard evidence for this interpretation, it presents an idea to look for in future work. We feel that if this is an actual tendency (try to read familiar plans even into novel situation), it has important implications for intelligence analysis. For example, if the enemy implements a novel plan that sends out unexpected indicators, analysts may still try to interpret the new plan as an old, familiar plan, rather than come up with novel interpretations. If so, blue forces would be susceptible to these types of operations.

Discussion

The project ahead technique offers two contributions to the knowledge elicitation process and understanding how experts use their knowledge. First, PA is a useful tool in determining how expert's structure problems. The scenarios they generate can serve as the basis by which experts determine what events/enemy actions they expect to see, which in turn could influence actions

taken by the experts. Therefore, understanding these scenarios and how the expert generates them is an important requirement of understanding expert problem solving.

Second, the project ahead can serve as a basis for setting up further elicitation. By eliciting scenarios or sequences of events that experts reason with, the elicitor generates a context for elicitation of such things as what evidence the expert's use, how they reason with information, what indicators they expect to see, etc. We discuss each of these two issues in turn.

With regard to the first issue, what scenarios experts generate and when they generate them.

Elicitation Using Conflict Resolution

Methods

Subjects. Subjects were eight enlisted personnel and officers recruited from the XVIII Airborne Corps, Fort Bragg and III Corps, Fort Hood, and their subordinate units. Subjects ranged in rank from sergeant to lieutenant colonel and represented specialties of situation development/all-source analysis and collection management.

Materials. The Conflict Resolution (CR) technique utilizes knowledge elicited from other techniques and explores that nowledge in greater depth, ferreting out underlying assumptions, nditions upon which the knowledge depends on, etc. Hence, CR typically follows problems solving or Cognitive Structure Analysis techniques. In cases where CR followed problem solving, the full-scale, context driven problem described in the problem solving techniques section above was used.

Procedure. Of the eight subjects, six were run in pairs of two subjects each (two pairs of situation development/all-source analysis and one pair involving an all-source analyst and a collection manager). The remaining two subjects were run individually. The all-source analyst who was paired with the collection manager was also run individually in a separate session. Hence, there were six conflict resolution sessions in all. Sessions lasted from three to eight hours although part of each session was devoted to problem solving and/or Cognitive Structure Analysis to elicit the knowledge that CR was to expand upon.

Results

Data Used. The data used here represent an amalgamation of the three sources of data: elicitor's notes, on-line transcription,

tape recording. Essentially, the elicitor went back over the on-line transcriptions and revised them based on the session notes and tape recording. These revisions focused on filling in gaps and correcting errors.

Once the transcripts are revised, they served as the primary source for extracting elicited knowledge.

Analysis. As discussed, the goal of the CR technique is to gain insight into how the expert reasons with data and to ferret out underlying assumptions that often remain implicit during interview and problem solving sessions. We illustrate the use of this technique by presenting excerpts from a session with a pair of Order of Battle experts. We then illustrate the types of knowledge structures and knowledge use rules that can be derived from this technique.

The text below is taken from the transcript of the sessions with the subject matter expert. It illustrates some of the probing sequences in CR. "I" denotes the interviewers queries, while "E1" denotes the first expert's responses and "E2" denotes the second expert's responses.

(During an earlier part of the session the subjects had been asked for indicators of the location of an attack.)

> E2: "Movement of the Division Command Post (CP) is an important indicator of where the enemy division is planning to attack."
>
> I: "What is the probability that the enemy division will move its command post into the region where it plans to attack?
>
> E1 and E2: "Absolutely certain. 100%"

(If in fact the experts were correct that movement of the command post (CP) is 100% diagnostic of the location of the enemy attack, then it would be a powerful indicator indeed. However, psychologists have identified an "overconfidence bias" (Lichtenstein, Fischhoff, & Phillips, 1982; Kadane & Lichtenstein, 1982) that often occurs in judgment tasks. Therefore, it is essential to know whether the indicator really is that strong or whether there are conditions that would reduce the diagnosticity of that indicator. The elicitor attempts to resolve this issue by creating a conflict for the subjects to resolve.)

> I: "Now, I'd like you to imagine that I have an infallible crystal ball and it tells me that the enemy division is planning to attack in the central region (the location of the attack is arbitrary for the

present purpose), but the command post will not be moved toward the central region. How would you account for that?"

E2: "The division command post could have been destroyed enroute, and the Army command post could have taken over its function." (Qualification-1)

E1: "To save time, and increase the chance of surprise, the division commander and his staff might not relocate their equipment, but instead they might occupy facilities already in place in the forward area." (Qualification-2)

I: "Now my infallible crystal ball tells me that the enemy division is planning to attack in the central region and the command post will not be moved to the central region; yet neither of your explanations is correct. Can you explain that?"

E2: "Time may be so critical that the division commander decides to attack without doing the normal doctrinal reconnaissance." (Qualification-3)

E1: "If time is critical, the division commander might delegate responsibility for recognizance to the rear command post of the first echelon division." (Qualification-4)

I: "The crystal ball tells me that none of these explanations are true. Any more explanations?"

E1: "The Army commander might have coordinated intelligence collection with the division commander and gotten all the intelligence he needed from air assets. They might both be airborne together." (Qualification-5)

E2: "The Army commander might have taken over the division from an airborne Army command post." (Qualification-6)

E1: "The Army commander could also take charge if there were a rail line in the area; he could use a train-borne command post." (Qualification-7)

For purposes of comparison with the earlier probability judgment, the subjects were asked for probabilities for each of the seven exception conditions, in the presented scenario (the Fulda Gap scenario). These were the results:

Qualification	E1's probability	E2's probability
1	.8	.8
2	.3	.3
3	.4	.65
4	.35	.35
5	.1	.3
6	.15	.15
7	.2	.03

Although there is a small amount of disagreement between the two subjects, the probabilities for various exception conditions are significant. Assuming the exception conditions to be independent, the probability that none of them is the case is .03 for the first expert (E1) and .02 for the second expert (E2). If these are all the exception conditions (and there is no reason to think they are), the strength of the relationship between moving the command post and location of enemy attack has been reduced from 100% to about 3%!

Discussion

Conflict Resolution is clearly a corrective for overconfidence. Yet it would be an equal mistake to suppose that the probabilities computed from the exception conditions are more valid than the original assessment of absolute certainty. With such low probabilities, intelligence analysts would never be able to draw any conclusions or make any recommendations at all. In the normal course of events, most if not all of these qualifications can be disregarded: the analyst may think and act as if the probability of the command post being moved were 100%. In case of the unexpected, however, they may become relevant: e.g., if some evidence suggests an attack will occur in the center, yet the divisional command post does not move there.

More important (for present purposes) is the effectiveness of Conflict Resolution as an elicitation tool. It starts with a rule that characterizes the automatic or stereotypical expectation (If the enemy division plans to attack in region X, then it will move the divisional CP into the region; $p = 1.0$), but leads to a much fuller understanding of the "theory" underlying the rule. We learn:

(1) The goal or reason for moving the CP forward (recognizance related to the planned attack);

(2) A competing goal (surprise) that might outweigh the primary goal in some circumstances;

(3) A variety of alternative means for achieving the goal, shedding light on both physical and organizational constraints:

Organizational--coordinating with the Army versus being taken over by the Army versus reporting to the Army;

Physical--using air versus rail versus a fixed site (relevant only with Army control or coordination);

Physical/organizational--moving personnel without equipment versus moving personnel with equipment versus utilizing the first-echelon command post's personnel and equipment (relevant only with division control);

(4) The possibility of failure and the command post's vulnerability to attack.

All of these topics can serve as starting points for further elicitation.

Where do exception conditions come from? It is possible though unlikely that they are directly associated in the subject's memory with the original rule, but not articulated except in a situation of conflict. A far more likely possibility is that they are generated during the session from more fundamental knowledge, which is typically activated only when a more superficial approach runs into trouble. Such knowledge identifies intermediate causal links between the antecedent and the consequent of the rule, and specifies how each link could break down. After eliciting a set of qualifications to a rule, therefore, the elicitor may try to model the knowledge by assigning each qualification to an appropriate intermediate causal step.

Figure 1 is a preliminary, tentative model for the inference in the present example. The *beginning* and end points of the inference are boxed: i.e., a decision to attack in region X is the antecedent, and the successful movement of personnel and equipment of the divisional CP into region X is the consequent. The bold arrows show the normal or expected course of events between these two points: the Army permits the division to retain responsibility for recognizance in its own area, the division decides to perform recognizance and to do so by moving both personnel and equipment, and finally the movement is successfully executed. Exception conditions, shown by broken arrows, represent alternative outcomes at each of these intermediate decision points. Thus, the Army might decide not to allocate responsibility to the division, but coordinate with it or take control of it, utilizing rail or air to perform the recognizance function (qualifications 5, 6, and 7); if the division does receive responsibility, it may decide to forego the recognizance function (qualification 3); if it does perform the function, it may decide to do so by moving only personnel or by taking over a forward CP (qualifications 2 and 4); and finally, if it does decide to move the CP, it may fail to do so successfully (qualification 1). The causal arrangement of these qualifications explains why one of them has the power to invalidate the original inference.

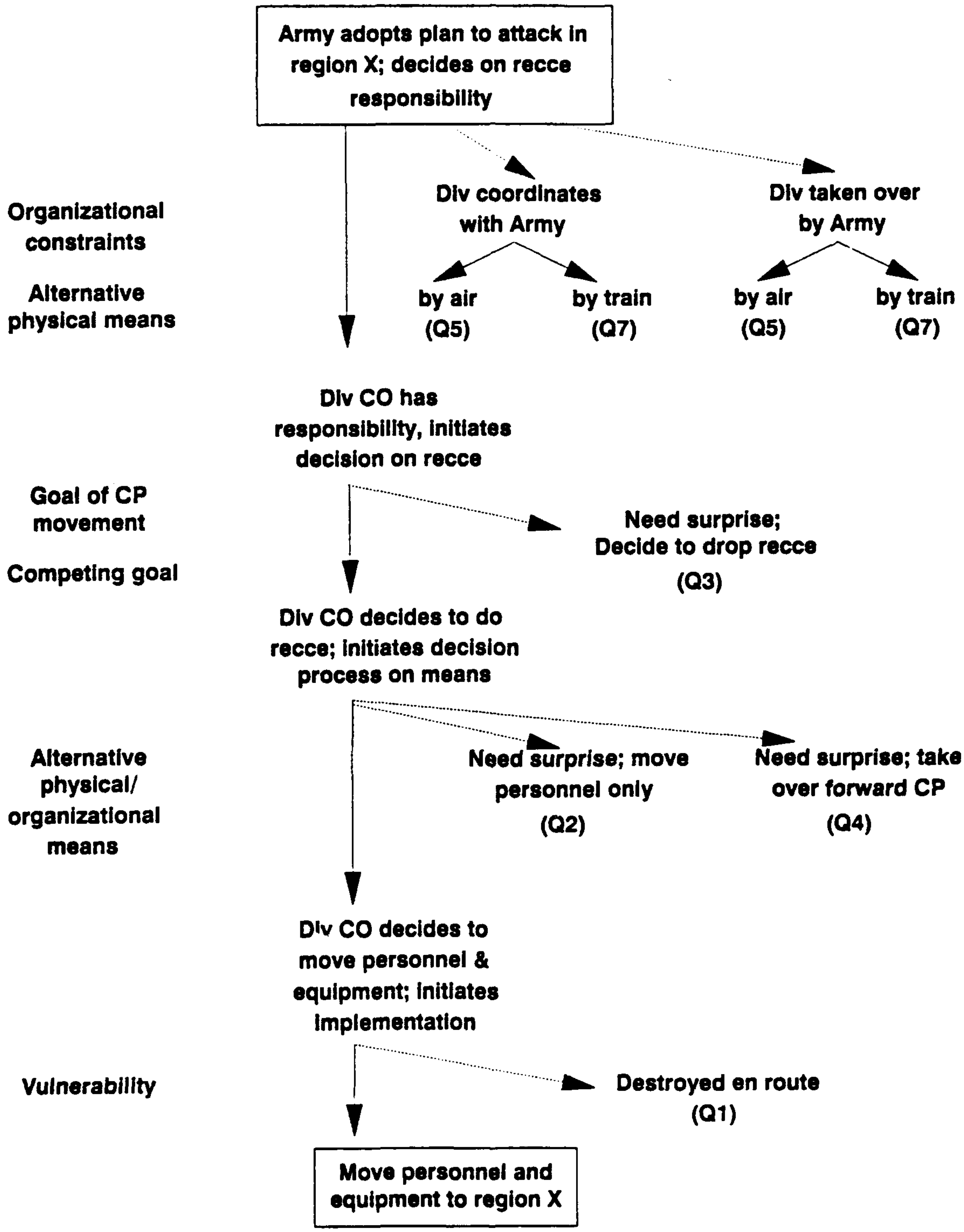

Figure 1. Theory behind a simple rule.

The causal model also sheds light on the subjects' cognitive processes. The qualifications were generated roughly in reverse causal order, suggesting that subjects began with the dictum of the crystal ball that the consequent was false, and worked backward in a search for explanations. The package of knowledge associated with each decision point in the sequence may have been accessed only when the previously activated package was shown by the crystal ball to have failed. This is compatible with a strategy of revising belief as little as possible in order to account for the conflict. For example, if qualification 1 explains why the consequent is false, then the entire causal chain up to the decision to move the command post remains as expected; if qualification 2 or 4 is invoked, however, the causal chain is intact only up to the decision to perform recognizance; qualification 3 means that only the assignment of responsibility to the division is preserved; and qualifications 5, 6, and 7 do not preserve even that. As the elicitation session proceeded, therefore, these subjects were pressed by the crystal ball to question increasingly extensive aspects of their normal response. (The ordering of qualifications 2, 3, and 4 suggests that for these subjects, the decisions to perform recognizance and to do it in a certain way were either accessed at the same time or actually represented as a single step.)

In this example, the subjects' initial response to the request for a numerical probability (100%) was provided with little hesitation. In other cases, however, subjects may find the request for a probability puzzling and difficult to comply with. CR, by prompting subjects to explore underlying knowledge, can clarify the nature of the required assessment. For example, during an early part of the same session, the subjects were asked the probability of a "false alarm," i.e., that the analyst would conclude that a CP had moved, while in fact it had not. Both subjects responded that they can't "assess this." Later, the subjects were asked to imagine (via the crystal ball) that the CP did not move, but that the analyst was convinced that it had moved, and to explain how that was possible. They found this "easy" and generated nine separate possibilities together with probabilities for each. The exception conditions covered a variety of important source of information regarding the location of the command post, together with various ways in which each of them could go wrong:

Intentional deception on communications (Q1)
Transcription error on communications (Q2)
Mishearing on communications (Q3)
Misinterpretation of photo (Q4)
Intentional deception on photo (Q5)
Human asset (e.g., stay-behind) misidentifies CP-type element (Q6)
Refugees misidentify site (Q7)

Internal communication is misunderstood (e.g., order to look for CP is taken as report that CP was found) (Q8)

CP movement inferred incorrectly from doctrinal template (Q9)

It is not surprising that direct probability elicitation was unsuccessful in this example; the antecedent of the statement may seem hopelessly vague. It is a highly aggregated description that applies to many different ways the analyst could become convinced of the CP's movement. For CR, however, its usefulness lies in its vagueness: the crystal ball forces the subject to desegregate the different information sources and methods of analysis, in order to specify the ways that the analyst could be mistaken. Such vague formulations might serve the elicitor in a top-down fashion, first to identify the various sources and methods, and then to apply the crystal ball to the more precisely specified relationships that emerge. After exploring various specific relationships, the elicitor might apply the crystal ball to a more aggregated statement in order to test for completeness.

CR elicits the conditions under which a relationship may not hold; the result, we have just seen, is typically more natural and more informative than asking for a numerical summary of the strength of the relationship. Why not, however, ask subjects to state those conditions directly? How important is the CR task *per se* for the success of this approach? Quite important, according to our informal comparisons of the two approaches. In several instances, the same subject in the same session was asked both (1) to specify conditions under which a relationship might not hold, and (2) to resolve a conflict in which the relationship did not hold. For example, two different subjects (than the ones discussed above) conclude that the second echelon division (in the problem solving scenario) would attack within the next 24 hours. CR, using the crystal ball technique, immediately elicited two exceptions: the division might not attack if they had trouble crossing rivers due to air attacks, or if logistics problems forces a delay. Subjects were then simply asked if there were any other reasons an attack might not occur or be delayed. Both subjects answered no. The crystal ball technique was then used again; the subjects were told that the division failed to attack within the designated time, but both of the reasons given earlier were false. In response, the subjects were able to generate six additional exception conditions. These included additional problems of executing an attack plan (e.g., equipment failure), but also raised more fundamental issues regarding higher-level goals and enemy doctrine (e.g., no penetration across front to exploit, enemy strength reduced below what is necessary to move forward, concern about blue air strength, enemy commander's personality) and about coordination with other units (enforced delay to allow other units to catch up). This session was typical of all subjects in that CR provided additional information after direct questioning had run dry.

CR is completely general in it applicability; it can be applied to any statement whatsoever made by a subject. The elicitor need only follow the line of questioning presented in Appendix A and to "consult the crystal ball" to find that the statement is false and ask for an explanation. The examples already given shed some light on the versatility of the method and the way in which it complements either a "top-down" or "bottom-up" elicitation strategy. In particular, CR can be applied to a general rule (e.g., If an enemy division plans to attack in region X, then it will move its CP into region X.), or to a specific instance of the rule (e.g., If the 9th Tank Division plans to attack in the center, then it will move its CP forward in the center region.). The elicitor may begin by applying the crystal ball to the general formulation, and then test the results by applying the crystal ball to various particular instances (specific divisions, specific regions) in a concrete problem. Alternatively, he might begin by challenging a particular conclusion adopted by the subject in a concrete problem, and then move on to test a more general formulation. Similarly, CR can apply to a relatively precise statement (e.g., If the unique call sign of a commander is detected in an area, then the commander is physically present in the area.) or to a relatively imprecise statement (e.g., If an analyst becomes convinced that the commander is physically present in an area, then he is present in an area.). Vagueness is especially useful in a problem solving scenario when the subject arrives at a conclusion (e.g., the 9th Tank Division will attack within 24 hours) with no explicit antecedent at all.

Another application of CR involves eliciting knowledge not about inferences, but about activities and procedures. In this case, a statement by the subject regarding what he would do under certain circumstances (or, in a problem solving setting, an actual action taken by the subject) is contradicted by the crystal ball. For example, in a discussion on intelligence collection strategies, two subjects stated that collection is always focused on the enemy's first echelon forces. The crystal ball, however, "showed" that a collection manager was not focusing on the first echelon, and the subjects were asked to explain how that could occur. A combination of CR and direct requests for exceptions generated five exception conditions. These exception conditions exposed the theory underlying the automatic rule of attending to first echelon forces, i.e., the goal of the procedure (to attend to the units with the greatest potential for immediate impact on us) and alternative means to achieve it in a variety of circumstances (depending on enemy actions and our own plans). In particular, collection assets might be focused on other echelon forces if the enemy is currently in a penetration attack so that close units are moving away from us, if the enemy is bringing up nuclear weapons in the rear, if the enemy is initiating a major action on our flank, if we are planning a deep attack, or if we are hoping to induce the enemy to try a penetration.

Elicitation Using Cognitive Structure Analysis

Methods

Subjects. Subjects were 17 officers recruited from I Corps, Fort Lewis and III Corps, Fort Hood and their subordinate units. Subjects ranged in rank from Warrant Officer First Class to Lieutenant Colonel and represented specialties of situation development, order of battle analysis, target development and collection management.

Materials. Even though Cognitive Structure Analysis does not require the use of any explicit materials, it is often useful to integrate the knowledge elicited from a given elicitation technique with the knowledge elicited from other techniques. Therefore, the scenario used in the other techniques described above was available as a reference for subjects as well as to define the scope of the knowledge to be elicited.

Procedures. Nine of the 17 subjects were run individually, while eight of the subjects were run in pairs. One subject was run twice, once as an all-source analysis expert (who is responsible for both order of battle analysis and situation development) and once as a target development expert. Hence, a total of thirteen sessions were run using Cognitive Structure Analysis. Each session lasted approximately three to four hours, depending upon the availability of the subjects.

The subject area of the elicitations varied according to the expertise of the subjects. Of the pairs of subjects run, two pair were made up of all-source analysis specialists, while two pair were made up of a collection management specialist and a target development specialist.

Results

Data Used. The data used here represent an amalgamation of the three sources of data: elicitor's notes, on-line transcription, tape recording. Essentially, the elicitor went back over the on-line transcriptions and revised them based on the session notes and tape recording. These revisions focused on filling in gaps and correcting errors.

Once the transcripts were revised, they served as the primary source for extracting elicited knowledge.

Analysis. As discussed, the goal of the Cognitive Structure Analysis technique is to gain insight into how the expert organizes the knowledge that he uses to solve problems and perform his duties. The heart of CSA is a set of probes that try to determine the relationships and level of abstractions of the

different pieces of knowledge that have been elicited. We illustrate the use of this technique by presenting excerpts from a session with an order of battle expert. We then illustrate the types of knowledge structures and knowledge use rules that can be derived from this technique. Since use of knowledge in inferential tasks such as order of battle analysis depends on the inferential value of that knowledge, we elicited information regarding knowledge diagnosticity as well as knowledge content.

We present two types of knowledge diagnosticity, standardness and distinctiveness (cf., Galambos, 1986). Standardness refers to the likelihood of witnessing an event given the presence of a context (or a knowledge structure). High standardness means that an event is very likely given a context (e.g., one always pays in a restaurant so paying is highly standard or probable). Distinctiveness refers to the likelihood of a context (or a knowledge structure) given an event. High distinctiveness means that given the event, the context is likely (e.g., given that one asks the hostess for a table, it is likely that this event is occurring in a restaurant). Hence, standardness relates to how likely one is to see an event, while distinctiveness relates to what that event means. Standardness and distinctiveness need not be correlated, e.g., one always pays in a restaurant (high standardness), but people pay in a lot of other places as well so that knowing that someone has just paid for something does not necessarily indicate the restaurant context (low distinctiveness). Logically, events that are highly standard (likely to be seen) and highly distinctive (strongly indicate a context, e.g., attack) would be the best indicators for an intelligence analyst. However, the analyst may not have the luxury to pick the indicators he wants to observe (given the enemy is trying to prevent the analyst from discovering enemy intentions)--hence, how the analyst trades of these dimensions is of utmost importance.

The text below is taken from the transcript of the sessions with the subject matter expert. It illustrates some of the probing sequences in CSA. "I" denotes the interviewers queries, while "E" denotes the expert's responses. After the elicitation sequence is illustrated, a representation of the elicited knowledge (which goes beyond the sample presented here) is presented along with some high-level reasoning/inference procedures that illustrate how the expert reasons with the different indicators elicited.

> I: "I'm also interested in certain sequence for these things, is there a compulsory sequence in terms of supplying ammo, missile support from FROGS, cross attaching" (Elicitor is taking the events the expert says comprise an enemy attack and is probing there logical sequence.)

E: "Supplies would come in 36 hours in advance. You need time to move out of supply dumps and distribute rations to troops. These things take time. The Soviets aren't any better than we are. For artillery support, they would try to move under cover of darkness or fog or low clouds. If there is no way, they would move through back roads through forest."

I: "Is the only constraint [for the artillery], that they get there by 1800?" [when the attack would kick off in the example used]

E: "Again the main constraint is to get there in time for the attack and not get shot doing it. Hiding is deception operations. Let's say they want to move battalions up. They would leave dummy guns in place and move through the back roads."

I: "What's the purpose of the artillery barrage?"

E: "The purpose is to inflict casualties to make us take cover, dig in and duck heads, and also blow holes in defenses so they can get through."

I: "So the objective is to facilitate that initial breakthrough?"

E: "Exactly. Like in WWI, the marines shelled that place for 2 or 3 days before they landed. In any war movie, before they attack islands, the planes and the Navy are always shelling first before they move in. If you can weaken them a little bit, it will help."

I: "Why would they use dummy guns to divert us?"

E: "We use imagery a lot from a fast moving airplane. A dummy gun we couldn't tell. They replace the real ones with dummies and move real ones. It diverts our attention and we waste ammo. If we waste $2500 on $2 worth of plywood, then it's not worth it."

Based on an analysis of the session with the expert, we developed the following representation of enemy attack operations:

MOP-ATTACK

Scene/Event: Air Mobile Operation

Reason: Hold key area until ground troops get there.

Timing:

Standardness: Medium probability.

Distinctiveness:

Scene/Event: Increased recognizance activity

Reason: To learn about changes in a fluid situation.

Timing: Throughout several days before attack.

Standardness: Fairly probable (depends on how well enemy already knows situation.

Distinctiveness: Moderately distinctive.

Scene/Event: Move up engineer assets (obstacle breaches, mine plows, sappers).

Reason: To facilitate movement of forces.

Timing: 24 hours before attack.

Standardness: High probability.

Distinctiveness: High distinctive (only two purposes for bulldozers: build or tear down obstacles).

Scene/Event: Remove obstacles (e.g., minefield).

Reason: Facilitate movement of forces.

Timing: Shortly before attack or little by little over several days.

Standardness: If obstacles in place, then very high probability.

Distinctiveness: Highly distinctive.

Scene Event: Prepare Air Defense (check equipment--radars, missile launches, guns).

Reason: If intending to fly something across FLOT, want to shoot down airborne defenders.

Timing: 24-36 hours ahead of attack.

Standardness: High probability.

Distinctiveness:

Scene/Event: Supply activity (ammo, parts, fuel).

Reason: To sustain attack.

Timing: 24-36 in advance of attack (depends on time required to get from supply depot to troops).

Standardness: Certain look at supplies they have and what they would need to determine what to expect. Often don't know they have --e.g., concealed in trees. May attack without adequate supplies for sake of speed.

Distinctiveness: Not distinctive unless localized to a particular unit or heavier than normal (e.g., 200%).

Scene/Event: Move up artillery support, emplacement of artillery.

Reason: Prepare artillery barrage.

Timing: To get there in time for attack. Use darkness, fog, back roads, deception to avoid being targeted. (Example of deception: leave dummy guns behind to direct friendly forces from real guns, to cause friendly forces to target dummy guns and waste ammo.)

Standardness: Certain.

Distinctiveness: Highly distinctive.

Scene/Event: Allocation of artillery units from Corps to division or from division to regiment.

Reason: Beef up forces for attack.

Timing:

Standardness: Highly probable from division not probable from Corps.

Distinctiveness: Strong indicator.

Scene/Event: Diversion.

Reason: To direct friendly forces from main attack.

Timing: 5 hours ahead of attack (since we need 4 hours to move our forces). (They will observe our movements in response.)

Standardness:

Distinctiveness:

Scene/Event: Back-up units moving forward.

Reason: To support attacking force.

Timing: 6-12 hours before attack.

Standardness: Very probably.

Distinctiveness: Routine if units have been in line a long time.

Stronger indicator if attack indicators are present:

- more compact
- looking for assembly area to group
- more heavily supplied
- higher speed
- high security profile

Scene/Event: Go into formation (2 up and 2 back).

Reason: All four units on line would be too thin.

Timing: 6 hours prior to attack, enough time to be in position 1-2 hours before attack.

Standardness: 60% probable. Won't attack as well as in a formation that was not practiced.

Distinctiveness: Fairly strong, although could be defense in depth or mobile defense.

Scene/Event: Cross-attaching extra tanks (among regiments).

Reason: To beef up attacking force.

Timing: Arrive 2 hours before attack--timing depends on distance.

Standardness: Lesser probability--Soviets don't do a lot of cross-attaching--but will do it if there is a strong Blue force.

Distinctiveness: Strong indicator.

Scene/Event: Massing of forces.

Reason: To increase chance of successful penetration of FLOT.

Timing: 3-6 hours before attack; will occur where we can't see them--heavily forested areas, at night.

Standardness: Highly probable.

Distinctiveness: Strong indicator.

Scene/Event: Missile support from FROG.

Reason: To weaken defending forces.

Timing: 2-3 hours before attack.

Standardness: Depends on supply, importance of attack.

Distinctiveness: Moderately distinctive.

Scene/Event: Radio silence.

Reason: Conceal movement, final positioning before attack.

Timing: Up to 1 hour before attack.

Standardness: high probability.

Distinctiveness:

Scene/Event: Artillery barrage.

Reason: Inflict casualties, make us take cover in order to blow hole in defenses so they can get through.

Timing: 1/2 hour prior to attack.

Standardness: may not occur if they have infiltrated people behind our lines or they want complete surprise.

Distinctiveness:

Scene/Event: Non-persistent chemical weapons.

Reason: To weaken defenses but not harm attackers.

Timing: First wipe out filters, then introduce blood agents, then nerve agents.

Standardness: High probability, especially if they've used them in past few weeks.

Distinctiveness: Moderately distinctive--Soviets consider chemical weapons as ordinary part of warfare, could use them for defense.

Scene/Event: Narrowing of sector.

Reason: To increase chance of successful penetration.

Timing: Time of attack.

Standardness: Very probable--depends on strength: stronger forces can handle wider frontages.

Distinctiveness: Very strong indicator.

Scene/Event: Close air support.

Reason: To weaken defenses.

Timing: 5 minutes before attack.

Standardness: High probability--Soviets not big on close air support except for recent use of helicopters.

Distinctiveness: Moderately distinctive (if there are other indicators) if lands and picks up troops, then highly distinctive of air mobile operations.

Scene/Event: Attack.

Reason: To achieve penetration.

Timing: Before dawn (if from east then sun in our eyes) at night (if our units are know to have failed night defense).

Standardness: Fairly high probability.

Distinctiveness:

This structure is not simply a compendium of inert facts. It may be used dynamically to generate a very large number of different inferences. In the absence of a coherent knowledge structure, these inferences would be explicitly described by many low-level rules. A very small number of general "meta-rules" of the following sort, however, may be all that is needed to generate these inferences in the presence of the knowledge structure.

1. Determining what indicators to look for.

(a) If an indicator of one event is observed, intensify search for indicators of other events.

(b) If an event would be highly distinctive of attack if it were observed, then give its indicators priority in the search (e.g., moving up artillery).

(c) But if an event has a very low probability, give its indicators low priority in the search (e.g., cross-attacking forces).

(d) If an event is distinctive of a type of attack for which defenses are poor (even though it is not distinctive of attack as such), and the event is reasonably probable, then give its indicators high probability in the search (e.g., chemical weapons).

(e) The stronger the observed indicators, the more search effort (i.e., special tasking) should be requested (e.g., bulldozers move forward).

2. Determining when to look, how frequently to request reports, how significance of an indicator degrades with time.

(a) If time of attack has been independently established, search for an indicator at the time it is expected to be observed in relation to the expected time of attack (e.g., look for resupply just after dark and at 2:00-6:00 in the morning, prior to established dawn time of attack).

(b) Request reports of indicators at a frequency determined by the expected time required for the function or goal of the events to be accomplished (e.g., if bulldozers are observed, report every 4-6 hours; if bulldozers are observed on the move, report every hours; if bulldozers are observed breaching obstacles, report at once).

(c) If the function or goal of an event is not attempted or accomplished within expected period, it becomes neutral as an indicator (e.g., artillery massed for 2 days).

Discussion

We view CSA as a valuable knowledge elicitation tool, both in its own right and as a supplement to other knowledge elicitation techniques. Also, we view CSA as playing a valuable role in expert system design, training and training evaluation, and personnel selection.

As a knowledge elicitation technique, CSA is designed to elicit both the structure and content of knowledge. Hence, it is a useful stand alone technique. However, CSA can also be used to supplement other knowledge elicitation techniques (see section below on integrating the separate techniques). Perhaps the most widely used knowledge elicitation technique is think aloud problem solving. This techniques provides rich sources of knowledge. Its major drawback is that the elicitor often cannot infer how the expert represents the knowledge being elicited or understand the process that gave rise to the output. Knowledge engineers have traditionally dodged this problem by focusing instead on how the system to be built will represent the knowledge and imposing that representation. However, we argue this practice severely limits the capability of the resulting system. CSA would be useful here since the elicitor could go back over the expert's output and elicit the underlying knowledge structures. Similarly, elicitation techniques which involve having experts make decisions (e.g., decision analysis, decision conferencing), make inferences (e.g., resolving conflict) and generate scenarios (e.g., wargaming) could also benefit by using CSA to uncover the structures which gave rise to the expert's output.

We feel that the results of the work we have done so far (cf., Leddo & Cohen, 1987; Leddo et al., 1988) have important implications for how expert knowledge can be applied and therefore how it should be elicited. Our work reveals that experts use all of the structures discussed in this paper (and perhaps others not yet defined) in an integrated fashion. We feel the issue of knowledge representation should drive the application of expert knowledge rather than get lost in the application. For example, if taken to heart, our work suggests that we should not think of rule-based vs frame-based expert systems, but rather systems which employ integrated representations. We are heartened that attention is turning toward systems which are both frame-based and rule-based, but note that such systems tend to be driven more by computational efficiency rather than compatibility with expert reasoning.

In the training arena, expert knowledge representations can serve as goal states which training can strive for and can provide criteria by which it can be evaluated. Here, CSA can play an important role in eliciting knowledge from students and comparing students' representations to experts' representations. Noting the differences can serve as a valuable tool for directing further training. To some extent, this is what happens in THINKER (Leddo, Sak, & Laskey, 1989), an Intelligent Tutoring System based on the INKS representations framework and which uses a simplified version of CSA to probe students' knowledge to assess what they have learned and how this compares to what experts know. We feel THINKER illustrates the promise that CSA has as a practical tool for diagnosing the structure and content of knowledge. Similarly,

personnel selection could be enhanced by identifying the types of knowledge experts have in a particular domain and using CSA to see if candidate applicants have the required knowledge.

Integrating the Individual Techniques

Each of the four techniques described above are designed to elicit different, but complementary types of knowledge and reasoning skills that the expert has. In addition to mapping out <u>what</u> the expert knows, the integration of these techniques attempts to map out the <u>constructive</u> processes the expert uses in reasoning with different types of knowledge as well as <u>how</u> the expert represents that knowledge (to the extent that such "representations" exist and can be approximated using current knowledge representation structures).

The key goal behind our integrated methodology then is to uncover the relationships and interactions between different types of knowledge and reasoning processes. For this reason, we argue that our integrated knowledge elicitation methodology is not a rigid, step-by-step sequence of queries, but rather a set of guidelines for eliciting and building links and relationships between knowledge. Hence, the elicitor is not presented with a decision tree of steps to follow, but rather a set of criteria for assessing the knowledge s/he has elicited so far and a set of guidelines for directing future queries. As a result, the individual elicitation techniques that comprise the integrated methodology are not seen as a set of discrete procedures to be performed sequentially, but rather as a set of knowledge elicitation approaches which are invoked according to the demands of the elicitation session and the evaluations of the elicitor. Hence, an elicitation session using our integrated elicitation methodology will typically be fluid, with the elicitor alternating between techniques as s/he sees fit.

In order to illustrate how our integrated knowledge elicitation methodology is implemented, we return to the basic goals of the elicitation process and our INKS knowledge representation framework (Leddo & Cohen, 1989; Leddo, Cardie, & Abelson, 1987). In knowledge elicitation, we are concerned with two primary objectives: (1) how do experts solve problems; and (2) how can the knowledge of how experts solve problems be captured in a way which is both psychologically valid (as close as we can get) and replicable in applications such as expert systems and training.

In assessing how experts solve problems, we are interested in several things. First, we are interested in the goals and objectives that experts pursue. This also includes the priorities and tradeoffs as well as compatibilities and inconsistencies among

goals. Next, we are interested in the planning and problem solving process that experts go through to achieve their goals. This complex body of knowledge is the heart of the elicitation process and encompasses many diverse types of knowledge. The planning process involves, among other things, a causal understanding of the situation that the problem solver is currently in and how it relates to his goals, procedures for generating plans to change the current situation to the desired one, actual "canned" plans which represent past steps that were taken in similar situations (which may be adapted according to the specifics of the current situation). We are also interested in the specific steps or procedures that the expert takes during planning and how the depend on the specifics of the planning situation. We are interested in the concepts and object knowledge that experts use since these represent the "things" (e.g., weapons, equipment) that expert work with. We are interested in the functions they serve and how they fit in the plans. Finally, we are interested in the causal reasoning and rationales that tie the whole problem solving process together. This includes why the plans lead to the goals, why specific procedures are instrumental to the chosen plans, and why the particular objects or concepts being used serve the functions they do. We feel the causal understanding of the planning process is especially important since this will enable us (or a resulting expert system) to replicate the expert or function in novel situations.

The elicitor, therefore, is guided by this general conceptualization of what he is trying to elicit. The elicitor must then be able to keep the knowledge elicitation session in perspective to get a sense of where s/he stands with respect to this goal (or any goal the elicitor has set up for him/herself). The elicitor then assesses what gaps or general directions s/he needs to proceed in. The elicitor must also understand how his repertoire of elicitation techniques relates to the knowledge s/he is trying to elicit. Finally, the elicitor picks the technique and proceeds with it according to the demands of the situation.

We offer some guidelines to the elicitor. In general, a good way to proceed is to get a general sense of how the expert reasons about the domain. This overview should cover basic things such as the expert's objectives, the kinds of situations s/he deals with, the kinds of people s/he interacts with, the equipment s/he uses, etc. This general orientation will help clarify a lot of what the elicitor will witness during the rest of the elicitation sessions. Generally speaking, we have found that CSA is a good technique to initiate the elicitation process to provide this overview. The elicitor should be attempting to sketch out the different types of knowledge outlined above (e.g., goal, planning) as this will serve as a valuable organizational tool for subsequent elicitation.

The responses typically elicited with CSA tend to be general. Hence, in order to assess how the expert really performs his/her job, we recommend a switch to problem solving, preferably with a full-scale, context-driven problem. The problem solving will probably form the heart of the elicitation session. As the elicitor articulates his problem solving process, the elicitor should continue fleshing out the sketch started during the CSA technique.

It is expected that most problems will involve some sort of information stream which the expert must process, whether it be background information (e.g., order of battle) or a message stream. The expert should be encourage to articulate how he is using the information. At this point, one might feel that it would be appropriate to switch to the Conflict Resolution technique which deals with information analysis. We feel that it is important to wait until the expert has spontaneously finished his piece of analysis first so as not to disrupt his natural problem solving flow. Our experience suggests that Conflict Resolution works best if we wait until the expert completes a piece of analysis (although he does not necessarily have to complete the whole problem) and then introduce the technique on those pieces of information which the expert states is most important (of course, the elicitor can apply the technique to whatever information s/he chooses). A good guideline for when to introduce the technique is to balance the demands of not interrupting the expert with the desire to elicit the knowledge while the analysis is still fresh in the expert's mind.

During the course of the problem solving, the expert may indicate that he is starting to project future events. This is likely to occur after the expert has become familiar with the problem (i.e., read the background information). Of course, the extent to which the expert projects these events will depend on the type of problem. This is a good time to introduce the Project Ahead technique since the expert is thinking in terms of developing the future events. We recommend using the technique while it happens. However, we recommend getting the expert to systematically project what events he foresees and then delving into the underlying reasoning or asking for specific details (e.g., "What will the artillery be doing as the front line units cross the line of departure?"). Again, these projections might be based on evidence they've seen (e.g., "Well, the lead unit is in the north so they will probably advance in the north along the autobahn."). When this happens, introducing the Conflict Resolution technique may be appropriate.

When the elicitor has completed the elicitation of the problem solving process, it is good to pause and assess where the elicitation stands (in general, regular breaks during elicitation sessions are useful, not only to give the expert a rest, but to

give the elicitor a chance to evaluate the progress of the session and plan future lines of query). Here, the elicitor should attempt to get a rough sketch of the knowledge elicited. If the elicitor has overnight, this will be helpful; otherwise the elicitor will have to do this during a break. This assessment will give the elicitor the opportunity to decide what loose ends to tie up and what types of knowledge structures to validate (below we describe how to analyze the elicitation sessions).

At this point, we recommend switching to the CSA technique to meet both objectives (fill in loose pieces of knowledge and attempt to validate structures). Keeping in mind that we noted that the CSA technique tends to lead to general commentary, it may be useful to have a set of mini-problems that the expert can work through to illustrate the specific points s/he's making. Of course, this requires some planning on the part of the elicitor to insure that a good supply of such problems is available (a military subject matter expert may help develop these). Finally, the elicitor analyzes the sessions and plans future ones.

Methods

Subjects. The subject used for this elicitation was a Chief Warrant Officer Third Class with more than twenty years active duty experience in the U.S. Army. The subject was an order of battle technician with the all source production section of the 1st Cavalry Division, Fort Hood. The subject was an expert in situation assessment and order of battle analysis. For the present study, he served as a situation assessment expert.

Materials. The present elicitation used a problem solving scenario. The scenario used in the sessions was set in the Fulda Gap region of West Germany. This scenario was well suited for the subject, whose unit had a European theater mission. The scenario depicted a U.S. division (the 52d MID) in a defensive posture against a Soviet Combined Arms Army (the 8th CAA), with a separate Tank Army (7th TA) as a potential follow-on force. The 8th CAA was comprised of three divisions, two relatively attrited Motorized Rifle Divisions (MRD's) in contact with the 52d MID and at nearly intact 9th Tank Division in reserve.

The scenario began with a lull in the fighting. The enemy had been conducting limited probing attacks and the reserve forces (9th Tank Division and 7th TA) had been advancing toward the FEBA (the boundary between friendly and enemy forces). The 9th Tank Division had settled into an assembly area approximately 15 kilometers (km) from the FEBA and the 7th TA had been continually advancing and its lead elements were approximately 50 - 60 km from the FEBA.

The subject was provided a variety of sources of information. In particular, there were 1:50,000 and 1:250,000 maps depicting division and Corps, respectively, areas of interest. For each of

the maps, there were overlays depicting friendly and enemy OB as of the morning of the current day's fighting. Since the subject was solving the problem from the division's perspective, most of the information provided related to things the division would be interested and to the 1:50,000 map. The subject was also given overlays of the terrain analysis of the division's sector and IPB products in the form of overlays depicting named areas of interest (NAIs), target areas of interest (TAIs) and an event template.

In addition to the graphic information, the subject had notebooks containing detailed information on enemy OB, a database of relevant events in the war, relevant Corps and division mission statements, weather and terrain information, and a summary of enemy capabilities and vulnerabilities. Finally, there was a message stream which presented processed intelligence information in the form of hourly INTSUMS.

Procedure. The knowledge elicitation was conducted over three sessions lasting approximately eight hours each (with breaks) over three consecutive days. The scope of the knowledge to be elicited was situation assessment, with emphasis on using enemy strategic and operational planning considerations on predicting tactical courses of action. The goal was to elicit knowledge that was more general to the European theater than just the particular scenario.

Prior to the actual elicitation, the subject was given a briefing on the background to the problem. This briefing was provided by a co-elicitor who had a military background. The co-elicitor was available to answer technical questions the subjects had about the scenario.

The sessions began with the problem, but the elicitor switched back and forth between techniques as needed. In the results section, excerpts of the sessions are shown to illustrate the criteria used when alternating between techniques.

All sessions were tape recorded. In addition an on-line transcription of the sessions was generated by an experienced typist. The elicitors also took notes of these sessions.

Results

Data Used. The data used here represents an amalgamation of the three sources of data: elicitor's notes, on-line transcription, tape recording. Essentially, the on-line transcription was revised based on the session notes and tape recording. These revisions focused on filling in gaps and correcting errors.

The revised transcript serves as the primary source for extracting elicited knowledge.

An Illustration of the Use of the Integrated Knowledge Elicitation Methodology

Below we illustrate how the integrated knowledge elicitation methodology can be used. We point out that the emphasis is not on showing how individual techniques are pursued (this is done in other reports of this series); rather our goal is to show how the individual techniques are used as part of a package. We reiterate that there is no fixed sequence of steps to follow. Rather, the elicitation session is an interactive process between elicitor and expert. However, the elicitor is aware of the types of knowledge and reasoning skills that he hopes to elicit and guides the experts along the lines that will give the most complete illustration of the expert's reasoning process and underlying knowledge. We point out how the elicitor shifts back and forth between techniques as he attempts to paint a picture of the expert's knowledge.

The text below is taken from the transcript of the sessions with the subject matter expert. "I" denotes the interviewers queries, while "E" denotes the expert's responses. After each "E", there will be a number that denotes the response number for this material (e.g., E1, E2). This will serve to identify what responses are being referred to when we build a knowledge representation later on. Comments in parentheses serve to explain the rationale behind the elicitor/interviewer's queries.

(The expert has been given the tactical problem used in these sessions. The elicitor decides to use a variant of the project ahead technique, where he asks the expert to project backward. Here the elicitor is interested in eliciting knowledge about how a present situation came to be as opposed to how it will unfold. The technique is essentially the same.)

> I: "Let's take the scenario back to day zero and explore how the situation might have evolved.... You have talked about a build up that might occur in a time frame of 2-3 weeks with massive preparation followed by a quick break. I'd like to discuss what that might look like, what indicators you would get, how would you know the build up was coming after first day, week etc."
>
> E1: "First, you have to consider before you can have war in Europe, the political situation overall has got to deteriorate. There has to be good reason for the Soviets to start war in Europe. Might be economically-related, based on 3d world conflicts, oil shortage, etc. It could be politically motivated by internally threatened break-up of Warsaw Pact, and weakening of Soviet grip on Warsaw Pact. Those are

probably the two most likely causes. What would happen is if it is economic problems, it would probably start somewhere in Middle East and the NATO build-up would a secondary thing. If break-up of Warsaw Pact the war itself would probably start within Europe. You would have warning that the Soviets are in trouble. The initial things you would see within intelligence community related to Europe would be the Comms (communications) Networks. With the linkage of comms initially from the General staff level down to front and to selected Armies and these comms would come up and be up for a while then they will go back down with a lot of stuff remaining in place. The next time they came up there would be additional stations netted into it and would be up and down and I would say that particular phase of it will start anywhere 30 to 60 days before invasion. A lot of that is typical exercise type build-up so it's not necessarily going to tip somebody off. The mobilization that would be required for Soviets to conduct sustained warfare in Europe would have to start taking place two weeks before actual hostilities commence. And the initial types we will see will probably be logistics personnel since they're so short on logistics. They have got so much stockpiled within the forward area, they would have to move out of fixed installations. We would start seeing massive movement of supplies out of fixed installations to field storage sites. Going back to comms, not only would the signal in CP types be out, but as we started drawing closer to date of hostilities we would expect to see exercises fairly significant ones whereas a typical exercise may involve up to elements of two separate Armies, we would see elements of all Armies as we got closer to hostilities dates; the exercises are more complex with additional forces added and they would start jelling towards which avenues the armies are going to use and start shifting toward that direction. One of the things we would expect to see within 1-2 weeks of hostilities would be departure of majority of their fleet from the Baltic Sea getting into the Atlantic, particularly submarines, to give them more maneuver room. Probably within a week of hostilities start at latest, we would have things such as departure of dependents from forward area. (I: What do you mean by dependents?) Soviet officers and senior NCO's, when they serve in Germany since they are there, senior officer tour can be anywhere from 1 year to up to 6-8 years. So they are like US army they have dependent housing these personnel, so their dependents would depart. The military liaison mission would be

restricted in its movements. The overflight collection would be curtailed due to closure of air corridors, etc. I would expect to see the Soviet military liaison become extremely active, particularly around such things as ammo supply points, particularly for special munitions, last minute targeting data, etc. "

(At this point the elicitor has choices. He can continue with the project ahead technique to flesh the events that would lead up to the war. On the other hand, the elicitor could pursue some of the loose ends left in the expert's response. First, the expert has discussed triggering conditions that would lead to hostility. Here, the elicitor can try to understand the causal relationships between the triggering conditions and the war's breakout. This would suggest switching to the CSA technique. Second, the expert has discussed some of the indicators he might see if these events were developing. Here, the elicitor could explore how the expert, as an intelligence analyst, would process these indicators as they were developing. This would suggesting switching to the conflict resolution technique. Third, the expert has been outlining a sequence of events which could suggest that he may have a prepackaged structure such as a script that describes this sequence. In order to verify whetner this is how the expert stores the sequence, the elicitor could switch to CSA and try to probe the structure. Realistically, the elicitor could pursue any of these paths. The elicitor opts not to pursue the third path right away since information elicited under the first two paths could further elucidate how the expert thinks about the problem. The elicitor chooses the first path first since he believes that eliciting the causal knowledge about why the war might start may be useful contextual information in understanding how the expert would handle the indicators as the preparation for warfare was going on. Hence, the elicitor switches to the Cognitive Structure Analysis technique.)

I: "Let's go back to two scenarios you outlined, i.e., economic scenario based on events in the Middle East and the political scenario based on the Warsaw Pact. Nothing is happening with the NATO relationship. It [the war breaking out] seemed almost independent of NATO relations. Unless that would be the underlying factor behind the other two. Does that fit into the equation?"

E2: "Not really, it [deterioration of relations with NATO] would be a symptom of the other two problems. For instance, as Warsaw Pact started to unravel and the Soviets and the hard-liners within each of the Warsaw Pact host countries started to crack down on dissidents and stuff along that line, part of what

they would have to do is crack down on journalists and blame a lot on outside interference on embassies and kick out diplomats and that would then cause increased tension or worsening relationships with NATO countries. But there is nothing that I foresee that the NATO countries could or would likely do that would upset the Soviets so much to actually start hostilities in the European continent."

I: "Conversely, you are saying that even if NATO fell into disarray that the Soviets have no incentive to attack unless they also have problems?"

E3: "That is correct. They can gain objectives much easier through political manipulation of various groups within NATO countries and attack on a NATO country, even if in disarray, would have the opposite effect--it would cause them to come together."

(The elicitor now decides to pursue the second branch in the directions discussed earlier that could be pursued after the discussion on events leading up to the war, namely how the expert would analyze the indicators as they were coming in. Therefore, the elicitor switches to the conflict resolution technique.)

I: "I would like to go back over indicators you have outlined, for example you started with comms network and how you would expect to see comms going up or down but you said that happens in exercises so it's tip off. My thought is if you could get that as tip-off you would be in great shape. So I want to try to compare what they might be doing differently in build-up to conflict vs exercises and see if you could get tip-off? What do they do different in war build-up or exercise build-up?"

E4: "Hate to say it, but nothing. You have to look at the whole reason for exercise build-ups particularly the way the Soviets have been doing it for past 10 years. Now it's designed to mask the build-up if they have to do it real-world. So they do everything they need to do to prepare for war. The intensity of it may be higher but that is such a subjective call that it would be difficult to convince anyone what that really meant."

I: "Since it's subjective, do you thing anyone would notice if were more intense?"

E5: "It would be noticed but those initial link-ups would be up for 3-4 days then back down and so by the time anyone would raise a red flag on initial hook

-ups and say, "hey something is different here", conflicting traffic [information] would come in."

I: "Are they trying to practice what communications would look like during a war? (The elicitor is trying to push the expert to think along different dimensions as to how the communications might be different in wartime vs peacetime.)"

E6: "Yeah, they are making sure they have good comms. When you look at Soviet comms network down near the bottom at the company and battalion [level]. Its fairly restrictive. Maybe within a company a there's a total of maximum maybe 2 nets, within a battalion maximum 3 nets, but when you start getting up into division, army, front, echelons above front their comms are extremely redundant. They are very paranoid about maintaining that strict command and control."

I: "Are they less paranoid at battalion and company?"

E7: "There's less effect at company or battalion. Normally a regiment will have a set mission--the regimental commander has briefed his company and battalion commanders and once the operation for the day has kicked off, it is fairly simple. If you have lost comms you know your mission is go to such and such a point and so other than stuff such as calling artillery fire your command and control is already set. When talking division level, the commander has to be ready within short notice to commit the 2nd echelon and reserves, to shift artillery, to accept additional artillery and air defense and engineers from higher, to be able to shift artillery from sector to sector, and his comms at that level are more important to him and higher he goes, the more important they become. Again looking at Soviets, their doctrine is based on operations level. So as long they keep the operational level side of the force going, then what happens at tactical level is less important."

(The elicitor now pursues the third branch, namely trying to flesh out the structure that the expert is using. Note, that the elicitor, by alternating techniques around a single theme, communications, is eliciting a wealth of knowledge regarding their role in the war and how the intelligence analyst interprets events regarding them.)

I: "From what you have said, I would infer that comms being set up is first to happen, why?"

E8: "Primarily because it is something you can do early. Take equipment up and set into place, take camouflage out and line of sight and wire in and leave it. It won't be that much of tip-off to have equipment off somewhere because when equipment is set up out there, after initial COMINT and majority comes back into garrison, then garrison comms come back up. It takes the I + W, the indicator of above normal comms [that would normally signal war was developing] and turns that off because now there is garrison comms in there. You can get up and out of way so when comes time to do last (2 week) massive activity to get ready for attack you don't have to worry about that because the comms are already set up. They're also there so they can control the last burst of activity."

I: "Are they using them during the two-week set-up?"

E9: "Yes."

I: "Is their use prerequisite to other activity? (This gets at whether there is fixed sequence dictated by causal relationships as opposed to merely setting up comms early due to convenience. The extent to which these causal relationships exist would support the idea of a script or MOP being present.)"

E10: "Yes."

(At this point the elicitor would go on to elicit the activities that depend on the communications being set up, i.e., the elicitor would continue with the CSA technique.)

Building Knowledge Representations of the Elicited Knowledge

In order to build a knowledge representation of the elicited knowledge, we follow the procedure outlined in Appendix A. Since the sample of knowledge presented above does not show extensive probing for its structural aspects, some inference will be needed to suggest an appropriate structure. However, we will state the basis for the conclusions regarding structure that we come up with.

First, we look for the major goals that might drive the outbreak of the war (which is the topic of the elicitation excerpts). In the beginning of response E1, the expert discusses two primary scenarios under which war might break out: (1) economic deterioration based on third world conflicts or oil shortages; and (2) break down of Soviet control of Warsaw Pact. While, the expert does not state explicitly that the Soviets have goals of economic and political stability, we can infer that these

goals are present and that they drive the onset of wars (the expert does corroborate the latter point). Since the expert did not explicitly state that the Soviets have these economic and political goals, we could seek corroboration through subsequent elicitation or gaining the opinion of a subject matter expert reviewer.

Next, we look for general plans that might unfold from these two goals. Again, in response E1, the expert states that the economic problems might lead to an attack in the Middle East with a secondary attack in Europe, while political problems might lead to an attack in Europe. Based on this, we can start building a knowledge representation that would look like this:

Goal:	Overcome Political Problems	Goal:	Overcome Economic Problems
Plan:	Attack Europe	Plan:	Attack Mid. East and Europe

Next, we look at the general procedures or events that might be involved in a European attack (this was the topic of the elicitation). Again, response E1 lays out the following steps:

Linkage of comms networks 30-60 days prior to attack.

Logistics supplies moving out of fixed installations to field storage sites.

Exercises with multiple Armies.

Departure of fleet from Baltic Sea to Atlantic (1-2 weeks prior to attack).

Departure of dependents from forward area (less than week prior to attack).

[Friendly] military liaison mission restricted in its movements.

Overflight collection curtailed due to closure of air corridors.

Soviet military liaison becomes active (e.g., around ammo supply points).

Once these major steps have been outlined, we look for details to flesh out these general steps. The elicitation mainly covered laying down of the comms networks (although some incidental details may have occurred in other places). Responses E1, E4, E5, E6, E8, E9, E10 bear on this issue. The structure gets fleshed out as follows:

Linkage of comms.

Comms laid out as in exercises, extra emphasis on higher echelons.

Come up for short while (3-4 days), then go down comms come up in greater number used in exercises.

Next, we look for concepts that were discussed in some detail. In this case, there were none in particular. Potential concepts that could have been covered were the nature of the communication networks. The elicitor chose not to get into this detail as this information was likely to be in manuals and was considered to be less interesting than the expert's reasoning process. The next step is to look for causal explanations of relationships between goals and plans and plans and procedures. We begin at the highest level of abstraction, namely, the relationship between the Soviet's goal of resolving conflict and the plan of attacking NATO forces.

The expert's responses E2 and E3 provide insight as to the reasons behind the attack: NATO would be used as a scapegoat as the Soviets cracked down on dissidents. This would lead to expulsion of NATO diplomats and worsening of relationships. The war would serve as a means to unite the Warsaw Pact.

Interestingly, there is a related explanation as to why the Soviets would not attack simply because NATO was vulnerable: they would attempt political manipulation because that would be easiest means to achieve their goals. Attacking NATO would serve to unify them, thus running counter to Soviet goals. This explanation was rather tangential to the topic being elicited though.

At the more detailed level, the expert offers an explanation as to why the Soviets are more concerned with communication link-ups at division and higher echelons. The following explanation is taken from responses E6 and E7:

> Regiments and below normally have fixed missions. If communications are lost, the unit can still carry out its mission. Divisions and above need to respond to instruction on short notice (e.g., commit reserve, shift artillery, accept assets from higher). Soviets emphasize this operational level of war [fought at higher echelons], while tactical level [fought at lower level echelons] is less important.

We can combine these various pieces of analysis of knowledge to form a unified representation. This representation not only will present a picture of the elicited knowledge so far, but can serve to guide future elicitations. The combined representation looks like this:

Goal: Overcome Political Problems

Rationale: NATO would be used as a scapegoat as the Soviets cracked down on dissidents. This would lead to expulsion of NATO diplomats and worsening of relationships. The war would serve as a means to unite the Warsaw Pact.

Plan: Attack Europe

Goal: Overcome Economic Problems

Plan: Attack Mid. East and Europe

Linkage of comms networks 30-60 days prior to attack.

Comms laid out as in exercises, extra emphasis on higher echelons.*

Come up for short while (3-4 days), then go down comms come up in greater number.

Used in exercises.

Logistics supplies moving out of fixed installations to field storage sites.

Exercises with multiple Armies.

Departure of fleet from Baltic Sea to Atlantic (1-2 weeks prior to attack).

Departure of dependents from forward area (less than week prior to attack).

[Friendly] military liaison mission restricted in its movements.

Overflight collection curtailed due to closure of air corridors.

Soviet military liaison becomes active (e.g., around ammo supply points).

*/Rationale: Regiments and below normally have fixed missions. If communications are lost, the unit can still carry out its mission. Divisions and above need to respond to instruction on short notice (e.g., commit reserve, shift artillery, accept assets from higher). Soviets emphasize this operational level of war [fought at higher echelons], while tactical level [fought at lower level echelons] is less important.

Once this general representation is laid out, we look for evidence that there is some knowledge structure or integrated set of structures that organizes this knowledge. In general, the goal and planning sequence of the initiation of war suggests a script or MOP. Some features of a MOP include a relatively fixed temporal sequence of events driven by causal relationships, plans that are linked to specific goals, specific triggering conditions related to goals. Based on the above representation, we see that the temporal sequence is largely fixed in terms of how events evolve as war approaches. There are clearly some causal relationships, e.g., the communications networks are used in the exercises (responses E9 and E10 confirm this). The attack plan is linked to specific political goals. In response E1, the expert states that the breakup of the Warsaw Pact would trigger the war.

Rationales are provided both at the highest level (i.e., why the attack on NATO would result from political turmoil) and at lower levels (why communication nets are more important at higher levels). Normally, we attribute mental model reasoning here. However, no explicit attempt was made to validate that such reasoning did occur. No real content knowledge was elicited, hence we represent no object frame or semantic knowledge. Finally, while specific procedures were listed under setting up communications, we cannot determine whether these are in fact production rules (i.e., are conditional on specific situations) or generic procedures that comprise a MOP or script.

Based on the representation presented, we can evaluate what further directions to take the elicitation. We note that the knowledge elicited is rather generic in nature, i.e., the knowledge describes scenarios of how the war might start that does not appear to be specific to any specific sequence of events. Here, the elicitor would determine, based on his/her goals, whether this level of generality was appropriate. In this case, our goal was to elicit more general knowledge of how the war might breakout. At this stage, the elicitor might decide whether to go after more context-specific (as in this case) or general knowledge.

As to other possible directions, we can delve deeper into the structural aspects of the knowledge and try to verify whether or not hypothesized MOP and mental models are those structures. This suggests using the CSA technique to cover the elicited knowledge. It may be interesting to see how the Blue forces are reacting to the Soviet build-up. Here, the Project Ahead technique would be useful. Finally, it might be useful to explore in greater depth how the intelligence analyst is reacting to events as they develop. For example, the expert stated that it is difficult to infer that the attack was developing based on setting up the communications links since the Soviets practice this in exercises. Here, the conflict resolution technique is useful to gain insight

into how the expert is reasoning about this data, as well as what other cues he's looking for (this in fact did occur later in the session).

Discussion

This section has described an integrated knowledge elicitation methodology driven by a theoretical framework of expert problem solving and knowledge use. This framework guides the knowledge elicitation process in two ways: first, it dictates what types of techniques need to be developed in order to capture different types of knowledge, and second, the framework generates a set of expectancies and frame of reference to evaluate the knowledge that has been elicited so that further elicitation can be directed. This latter point was illustrated in the excerpts of the knowledge elicitation framework.

While no explicit comparison was made between this integrated elicitation methodology and a standard, think aloud protocol technique, by looking at the flow of the session, we can argue that the integrated methodology enabled the elicitor to accomplish some things not otherwise possible by the think aloud method. For example, the elicitor was able to probe the expert regarding how he would interpret seeing the communications lines come up during wartime preparation. By using the conflict resolution technique (which tries to discover when information is diagnostic and when it is not), the elicitor found that the cue of communication nets coming up is not particularly diagnostic because it occurs that way in exercises. Had the elicitor not applied the conflict resolution technique, he may have inferred that since communication nets coming up was a step in wartime preparation, that such an observation was a strong indicator of imminent war. Such a conclusion could be dangerous if automatically encoded in an expert system.

This is not to say that such knowledge is never elicited using single knowledge elicitation techniques that are applications driven rather than expert driven. We are saying that it is necessary to dig further into the expert's reasoning in order to understand fully the intent behind the expert's utterances and to understand how the expert will use information in problem solving. Further, if we are to replicate the expert's behavior (as in an expert system), then it is imperative to understand how s/he reasons as well as merely observing the products of his reasoning.

By advocating a series of integrated techniques, we are acknowledging that experts (as well as other people) have diverse knowledge and reasoning skills. As a result, we do not expect a single elicitation technique to be "all things to all people."

Similarly, we do not believe that our integrated elicitation methodology has exhausted the range of reasoning skills and

knowledge patterns that experts have. Further development of this methodology would focus on identifying new reasoning skills and techniques to elicit that knowledge. These skills and associated knowledge structures would then have to be integrated into the INKS representation framework, which in turn would drive how the resulting techniques would fit within the integrated knowledge elicitation methodology.

We offer some ideas, based on our observations, of some interesting areas in which further work on knowledge representation and knowledge elicitation might go. While many experts rely heavily on "canned", doctrinal plans, there are occasions where the expert must generate new plans. While a lot of research has been done on the planning process (cf., Wilensky, 1983) that covers such things as planning kernels, etc., we feel the topic of how people select an appropriate planning strategy and then move from a general concept of a plan to a detailed plan needs further attention.

Clearly, in these situations a lot of pattern recognition and mental wargaming (simulation) is going on. As a result, the data-driven problem solving technique and the project ahead technique may be useful. However, there is undoubtedly a lot of causal reasoning which may not be getting captured. Perhaps a technique that might be useful here would be an adversarial wargaming exercise. This could be done by having two experts play opposing sides in a scenario (as opposed to most standard problems where there is no adversary). Here, the expert must carefully consider the consequences of his actions given that there is an adversary trying to exploit those actions. This may force the expert to make his causal reasoning more explicit. Research by Leddo and Govedich (1986) suggests that such a manipulation will actually induce a dramatically different reasoning style in problem solvers.

In the Leddo and Govedich study, subjects (who were not experts) were asked to plan versus implement military operations to rescue hostages taken by terrorists. Those asked to plan the operation (Planners) were in essence comparable to our present expert who approaches the problem as a "paper and pencil" task. Leddo and Govedich found that Planners approached the problem from an internal, causal consistency perspective (analogous to Tversky and Kahneman's, 1983, "intuitive" mode of reasoning). Planners failed to consider the validity of conclusions they were making, planned single courses of action with little in the way of alternatives, backups or contingencies--yet paid little attention to external events that could cause their plan to fail, and generally preferred complexity over simplicity.

Subjects who were asked to implement plans (Implementers) showed an opposite reasoning style, using an external validity perspective (analogous to Tversky and Kahneman's, 1983,

"extensional" mode of reasoning). Here, the "adversary" was a computer program which served as a simulation of the environment, including the enemy force, that the Implementers would face. Implementers did challenge the validity of conclusions that were made, liked to have alternatives, backups, and contingencies, did consider external events that could interfere with the plan's success, and preferred simple, rather than complex plans.

The conclusion here is that the adversarial process may cause the problem solver to examine and articulate his/her problem solving process. As a result, the underlying causal planning mechanisms may be elicited.

Comparison of Individual Techniques

We have presented a variety of individual knowledge elicitation techniques and argued that collectively they formed a highly useful, powerful knowledge elicitation methodology. Still there are basic questions that could remain regarding these techniques, e.g., "Do these techniques really complement each other or are they really redundant?" Also, given that the knowledge elicitor may have limited time with the expert, is there some way to maximize the amount of knowledge elicited for that given time, i.e., is one technique more efficient than another?

We tried to address these issues empirically. Rather than compare each of the four individual techniques, we dichotomized them into two categories: problem solving and interview (question and answer). These questions were then addressed in knowledge elicitation sessions conducted at Fort Bragg. We hypothesized that since interview techniques are more direct in eliciting responses from experts than problem solving techniques (where the subject must take the initiative to present his knowledge), the interview technique would be more efficient in eliciting knowledge (i.e., defined in terms of the amount of knowledge elicited per unit of session time). Further, we hypothesized that the knowledge elicited by the two techniques would be complementary (have little overlap). We argue that interviews, which tend to be conducted at a general level, access high level knowledge that is not organized by specific situations and contexts (i.e., require little priming to access). Problem solving tends to be very context-driven and the knowledge used tends to be more closely linked to specific situations and contexts. It is true that interviews can lead to specific examples and problem solving can lead to applications of general plans and knowledge--hence, some overlap between the two techniques can be expected.

Methods

Subjects. Subjects were five active duty army personnel recruited from the XVIII Airborne Corps, Fort Bragg. Subjects ranged in

rank from sergeant to major. All subjects were intelligence specialists with the exception of a captain who was recruited from the G-3 shop. This captain was run with the sergeant, who was an intelligence specialist.

Materials Used. The scenario used was the same context-driven, full-scale problem described earlier in the problem solving session. Only the standard problem was used.

Procedure. All subjects were run individually, with the exception of the captain/sergeant group described above. The knowledge elicitation sessions lasted a full day each (8 hours with a lunch break). The sessions were divided into two major components: testing the interview/problem solving technique tandem and testing the Personalized Decision Analysis technique (see Brown, 1989). Within each session, approximately four hours was devoted to problem solving and one hour to interview. The order in which these two techniques were administered was counterbalanced.

Results

Transcripts were generated of the individual knowledge elicitation sessions. These transcripts were then analyzed and their content used to develop INKS representations (see Appendix A for a description of how to do this) for the knowledge elicited for each subject for each of the two techniques. These INKS representations were then analyzed to obtain measures of the numbers of different types of knowledge structures (scripts/scenes, production rules, object frames, semantic net concepts, i.e., concepts with attribute linkages) and mental models) elicited per hour (hence the problem solving knowledge was divided by four since it was elicited over a four hour period). The knowledge structures were also analyzed to determine the extent to which different techniques elicited complementary, duplicative and contradictory knowledge. Complementary knowledge was defined as knowledge of a specific type (e.g., scripts) elicited using one of the techniques which did not duplicate other knowledge elicited from the same subject using the other technique. Duplicative knowledge was defined as the same knowledge elicited from the same subject using both techniques. Contradictory knowledge was defined as two pieces of logically inconsistent knowledge elicited from the same subject using different techniques. For example, if during the interview a subject stated that artillery was placed forward during the attack, then complementary knowledge might be that air defense if placed forward during the attack, duplicative knowledge would be a repetition of the statement regarding artillery, and contradictory knowledge might be a statement that artillery is placed back during the attack.

Technique Efficiency

Table 1 presents the mean number of different types of knowledge structures elicited per hour across subjects using the interview and problem solving techniques.

In general, it appeared that the interview technique was more efficient, eliciting over twice as much knowledge per unit of time as the problem solving technique. This would make sense, given in the interview technique, the elicitor is continually asking for responses from the expert, while during problem solving, the expert spends a lot of time working silently. Unfortunately, due to the small number of subjects used, the resulting analysis of variance (ANOVA) was not statistically significant, $F(1,3) = 2.82$, $p = .19$.

Interestingly, the interview technique did not appear to elicit more knowledge of every type than did the problem solving technique. The means for scripts and scenes are virtually identical for both techniques while the biggest discrepancies between the techniques appear to lie in the lower level knowledge such as production rules and semantic concepts. In fact, the overall technique by knowledge type interaction was statistically significant, in spite of the small number of subjects, $F(4,12) = 3.92$, $p = .03$.

Table 1

Mean Number of Knowledge Structures Elicited Using Interview and Problem Solving Techniques

	INTERVIEW	PROBLEM SOLVING	STRUCTURE MEANS
Scripts/Scenes	9.50	9.75	9.63
Production Rules	21.50	8.00	14.75
Frames	5.50	2.00	3.75
Semantic Concepts	11.75	3.75	7.75
Mental Models	8.75	3.5	6.13
Technique Means	11.40	5.40	

Informally, we had reported that in problem solving techniques, experts appeared to gloss over details and solve problems at a general level (see Leddo et al., 1988). Here we test this hypothesis statistically by performing a comparison between the number of high-level structures versus low-level structures elicited by each of the two techniques. High-level structures include scripts/scenes and frames while low-level structures include production rules and semantic concepts. Mental models are not included in the comparison because in our knowledge representation framework, mental models can be used to represent both high- and low-level knowledge. Table 2 presents the mean number of high-level and low-level knowledge structures elicited per hour across subjects.

In spite of the small sample size, this interaction came out significant, $F(1,3) = 9.9$, $p = .05$. This suggests that relative to problem solving techniques, interview techniques may be more efficient at eliciting low-level compared to high-level knowledge. At first blush, this may seem to contradict our original contention that problem solving techniques are valuable at eliciting low-level procedural and content knowledge due to the fact that problems provide specific contexts and primes for accessing this knowledge, while interviews activate general knowledge. While, we have no evidence to support our original contention, we note that the present study did not permit sufficient time with subjects to "saturate" their knowledge. Perhaps if, as in the case of expert system building, we exposed a group of experts to weeks of elicitation using interviews and another group of experts to weeks of elicitation using problem solving, the former group would saturate their low-level knowledge quicker than the latter. Unfortunately, such a study would have required access to lengths of time we were unable to get with our experts.

Complementarily, Duplicativeness, and Contradictions in Elicited Knowledge

In the INKS framework, scripts serve as the primary organizers of knowledge with scenes, production rules, frames, semantic concepts and mental models filling out the details. The elicitations were conducted so as to elicit these type of INKS representations (i.e., flesh out these different types of knowledge). Hence, the analysis of whether the two techniques are complementary, duplicative or contradictory depends on what knowledge they are going after. Here, two main topics were focused on the terrain factors of observations and fields of fire, concealment and cover obstacles, key terrain avenues of approach (OCOKA) and analysis of enemy courses of action. Within these two topics, both techniques were applied. The intent was that the two techniques should complement each other (i.e., this is the most desirable case for the sake of elicitation efficiency), hence the

Table 2

Mean Number of High- and Low-Level Knowledge Structures Elicited Using Interview and Problem Solving Techniques

	INTERVIEW	PROBLEM SOLVING
High-Level	7.50	5.88
Low-Level	16.63	5.88

ensuing analysis is primarily a measure of how efficient these techniques are in doing so. The analysis is broken down by type of knowledge structure within each INKS structure. Table 3 presents the mean percentage, across subjects and INKS structures, of complementary, duplicative, and contradictory elicited knowledge by type of knowledge structure.

As can be seen from Table 3, by and large, the problem solving and interview techniques can work in a highly complementary fashion in order to produce fleshed out INKS knowledge structures. There is little duplication in the elicited knowledge and internal contractions within subjects are virtually non-existent.

Problem-Solving Strategies Used by Subjects in Different Functional Areas

From the elicited knowledge, we are able to characterize the different types of problem-solving strategies used by subjects in the different domains. This serves two purposes. First, we can gain insight as to how different task and experience factors influence how experts solve problems. This could aid the development of expert and decision support systems. In addition, understanding the factors that drive the types of knowledge that are most functional in a given domain would be useful as a basis for developing tests for identifying people with suitable skills and problem-solving styles for specific jobs.

Second, by studying the different problem solving strategies of experts at different experience levels, we can gain insight into how expertise is acquired. This is an area that is not yet well understood. However, the implications for training are enormous. Novices could then be guided along a developmental path to induce them to represent and use knowledge as experts do.

The knowledge structures in the INKS framework are used to characterize the problem solving strategies of the subjects used during the project. For each type of knowledge structure, there exists a problem solving strategy that is based on it. We define each of these strategies.

A script- or MOP-based strategy is driven by goals and plans. This goal/plan strategy is defined as the process of selecting goals or objectives to meet and then developing action sequences designed to accomplish these goals. This can be viewed as a top-down strategy.

A frame-based strategy is defined as a template matching strategy. Here, people use the templates or patterns that comprise frames as a set of expectancies to pattern the environment after. People look for elements in the environment that fit the template (or discriminate between templates). Once the person has satisfied himself that the environment matches the

Table 3

Mean Percentage of Complementarily, Duplicativeness, and Contradictions in Knowledge Elicited Using Interview and Problem Solving Techniques

	COMPLEMENTARY	DUPLICATIVE	CONTRADICTORY
Scripts/scenes	89.20	10.8	0
Production Rules	94.33	4.83	.83
Frames	93.67	6.33	0
Semantic concepts	94.67	5.33	0
Mental Models	98.00	2.0	0

template (which is usually done on the basis of partial information), he uses the remainder of the frame to infer whatever information is missing.

A semantic-based strategy is similar to a frame-based strategy. The difference is that the semantic-based strategy is defined as processing based on semantic categories rather than frames. Here, problem solving is driven by looking at members of a particular semantic category (see Target Development below) or for semantic properties of an object.

A production rule-based strategy is defined as a process of generating procedures based on environmental conditions. These procedures can be used to generate action sequences or generate inferences. The main difference between the production rule strategy and the goal/plan strategy is that the former is driven more by environmental conditions (i.e., is more bottom-up), while the latter is driven more by goals (i.e., is more top-down).

A mental model-based strategy is defined as a process of causally analyzing events with respect to learning the outcomes that get produced. The mental model strategy is situation-specific and is often driven by goals (i.e., the person using this strategy cares about the outcomes as they relate to particular goals). The mental models themselves do not generate plans (hence, this strategy is not the same as the goal/plan strategy), but rather project outcomes of plans or event sequences. Hence, the mental model strategy can be used to support the generation or evaluation of plans.

Below, we characterize the different types of strategies used by subjects in the functional areas of collection management, target development, order of battle analysis, and situation development. We discuss variations in this knowledge across subjects, combat units (where appropriate), and potential strengths and weaknesses associated with each knowledge use strategy.

It should be noted that the knowledge summarized below was collected over a two year period at installations at Forts Huachuca, Carson, Bragg, Lewis, and Hood, and at sessions held at Decision Science Consortium (DSC). Subjects ranged in rank from sergeant to major general. In all, some 50 subjects were used. The knowledge focused on during the elicitation sessions was tactical intelligence, primarily at division and Corps level, against a Soviet threat in a European theater. A variety of techniques were used. These were primarily interview-based (i.e., question and answer) or problem solving-oriented. In the case of the problem solving, two basic scenarios were used: an offensive and a defensive "Fulda Gap" scenario. In addition, for the defensive scenario, a version containing a novel situation was

constructed and tested at Fort Lewis. Several "mini-problems", one for each domain was developed for in-depth probing of a specific process used as part of the general functional specialty (e.g., computing combat power ratios).

Collection Management

We observed two basic strategies for performing the allocation of collection assets. The first appeared to be more frame- or template-based. Here, assignments of collection assets seemed to be rather fixed, e.g., radars were assigned to brigades, LRPs (long range patrols) were placed deep along key avenues of approach (at named areas of interest or NAI's) and side looking airborne radar (SLAR) was instructed to fly up and down the FLOT to pick up whatever information they could. Subjects viewed this approach as a "vacuum cleaner," i.e., collection assets try to soak up as much information as they could. Interestingly, all subjects acknowledged that one can never collect all the information they needed. Yet, subjects adopting this approach did not seem to provide for prioritizing their information search. Hence, subjects using this strategy did not appear to address the vagaries of the particular situation they were in.

A second approach was more plan oriented. It was difficult to distinguish how much of these plans were MOPs (i.e., relatively preset) vs mental models (i.e., generated to meet the specific demands of the situation). Our best guess is that these plans were a combination of both--subjects appeared to use familiar plans and then "tweak them" in order to maximize their effectiveness in the present situation.

This plan-based approach to collection was driven by the commander's PIRs (subjects also discussed what PIRs they would recommend to the commander). Here, subjects would take the first PIR and decide how to best collect against it. This process involved breaking the PIRs down into specific indicators (i.e., observable events, activities, etc.) and deciding what assets they had or had access to (through other units) that could observe those indicators. The assets were then assigned to collect that information. The subjects then iterated the process to the next PIR. This process continued until either the subject ran out of collection assets or ran out of PIRs and IRs.

Once the subjects had completed this first pass at the collection plan, they began the "tweaking" process. Essentially, subjects looked for ways to leverage their collection process. This primarily involved looking for cases where a collection asset could collect on more than one PIR or could provide confirmation of another asset's collection findings.

In cases where subjects recommended PIRs, there tended to be a strong temporal/spatial component. Subjects seemed to follow a rule of "the closer the item is, the more important it is for collection." Hence, the units in front of the friendly unit were most important, followed by second echelon units, etc. This was qualified by whether or not the units in question could impact the friendly mission (e.g., nearby units that could not impact the mission were not important, whereas far away units that could impact the mission were important).

One aspect of this process that was conspicuous by its absence was that subjects using this strategy did not appear to make tradeoffs. Subjects did not try to maximize the <u>total</u> information the collected, but rather the total information on the current PIR/IR given the assets remaining after collecting against the previous PIR/IR. Hence, subjects implicitly acted as if the current PIR was more important than the remaining PIRs collectively. If this were indeed the case, then their strategy would be adaptive. However, if this were not the case, the subjects' strategy runs the risk of not optimizing the total information that can be collected. This could be especially important in cases where the top PIRs turn out to be the ones that are not the most relevant to the battle.

These two strategies did not appear to be associated with any specific unit. Rather, the type of strategy seemed to correlate with overall expertise. Subjects that were more "expert" appeared to use the plan-based method, while less "expert" subjects appeared to use the template-based. Here, our evaluations of "expertise" are rather soft. Primarily, our military consultants would give their impressions after each session of the quality of the material elicited and the level of expertise of the subject. These judgments serve as the basis for the above observation.

Each of these two strategies appear to have complementary strengths and weaknesses. The template-based approach is more balanced in that is tries to collect "some of everything." It would appear that this type of approach would show little risk of leaving large gaps in the collection process, although it may not provide the level of detail required for effective operations and may not respond well to the specific needs of the situation.

The plan-based approach focuses more on the perceived most important needs of the situation. However, as discussed above, this does so at the cost of a complete analysis of the battlefield (which may still be reasonable given the inability to collect all the desired information). However, this approach is more susceptible to potential cognitive biases on the part of the collection manager and intelligence analysts. For example, Tolcott et al. (1989) found that when subjects in the role of collection manager believed that an attack was to occur in a

specific location, they focused their collection efforts on that location, sacrificing other areas of the battlefield that could either disconfirm their impression or indicate that other threats were present as well.

Target Development

There were two primary approaches to target development. One could be viewed as a combination of templating/semantic and production rules. The other could be viewed as a mental model-based approach.

In essence, the first strategy employs specific target categories and prioritized targets by category membership. For example, enemy artillery might be the highest target priority, while enemy command and control might be next highest, followed by enemy maneuver units. Once these categories have been established, there may be rules that prioritize targets within the category. For example, for artillery, rules include the longer the range of the weapon the higher its priority and the larger the shell size the higher its priority.

The second strategy is centered around the concept of "center of gravity." Here, the expert tries to assess the enemy's "the center of gravity," i.e., the asset or part of the enemy that if destroyed, would cause the enemy's collapse. Examples of center of gravity include key political entities, key units or even key military leaders. In essence, the expert tries to figure out what makes the enemy "work" and then disrupt that process. The "center of gravity" concept can apply to different units at different echelons. For example, a division target development specialist is unlikely to be concerned with political centers of gravity (e.g., Moscow), but rather tactical and operational ones (e.g., Army command post).

Targeting centers of gravity go beyond merely identifying them. The expert must know how the enemy entity operates and the center of gravity's role in those operations. For example, at the operational level, the Soviet Army command post is the planning center that generates tactical plans for division and below. Once the plan is disseminated, the tactical units implement the plans with little need for guidance on their part (i.e., the plans remain relatively fixed). Hence, even though the enemy Army CP is the operational center of gravity that a friendly division may focus on, targeting that unit once the plans are disseminated may do little good. Hence, the friendly targeteer must both disrupt the planning cell (the Army CP) and insure that the planning cell is still required. This can be done by disrupting the current enemy course of action and then preventing enemy units from making contact with the Army CP (either by destroying it or jamming it).

It is clear from the above discussion, that this strategy for target development requires a causal understanding of how the enemy engages in warfighting and how it relates to the targeteer's unit. This causal understanding then drives the planning process which in turn arrives at a specific plan or set of procedures. This type of planning clearly illustrates our INKS concept. First, the targeteer starts with his goals (e.g., his mission). He performs a causal analysis to understand how the enemy can impact his goals. This in turn leads to a plan to produce a favorable outcome. The planning process is driven by a causal understanding of how the enemy operates in the domain and how these operations relate to friendly goals.

These two strategies seemed to be associated with specific units rather than specific people. For example, the first strategy (the semantic/rule-based approach) seemed to represent the target development strategy at III Corps, Fort Hood and its subordinate units, whereas the mental model-based approach seemed to represent the strategy at I Corps, Fort Lewis and its subordinate units. It is difficult to infer what gave rise to this organizational difference. On the one hand, it is widely claimed that the personality of the commander permeates a combat unit and the way it approaches warfighting. Hence, these strategies could have developed based on the leadership at these units. On the other hand, the two units fight in different theaters, on different terrain and against different enemies. III Corps fights in Germany against the Soviets, while I Corps fights in Korea. Perhaps the fact that Soviet doctrine has been so well studied that those confronting the Soviets have translated their mental models into concrete procedures and categories (cf., Cohen, 1989). If Korean doctrine is less well understood, then perhaps fighting against the Koreans cannot be reduced into concrete procedures but rather strategies for generating procedures. A potential problem with this interpretation is that both groups of subjects (Fort Lewis and Fort Hood) were interviewed on the European theater. Hence, if there were Army-wide procedures for fighting the Soviets, then presumably the Fort Lewis subjects should know these as well. One the other hand, the Fort Hood may have developed these procedures through intensive analysis of the Soviet threat, analysis going beyond standard doctrine.

It would be interesting to see whether other units with European theater missions do target development the same way as the units at Fort Hood. If so, this would support our explanation of the observed difference. In addition, it may suggest new research opportunities for understanding how knowledge structures evolve, e.g., from mental models to production rules/categories.

Again, each of these two strategies have associated strengths and weaknesses. If, in fact, the category/rule-based strategy represents a "compiled" plan for targeting (i.e., has routinized a

more complex set of analyses into simple categories and decision rules), then it gains speed and efficiency for a wide range of targeting scenarios. Such a strategy may be particularly suited for time stress situations. The drawback with such standardized approaches is that they fail to respond to the nuances of particular situations. However, such strategies may be at least "part right" for a large number of problems.

The "mental model" approach attempts to respond to the specifics of individual situations. As a result, they are more time intensive and may be less suitable for conditions of time stress. On the other hand, the mental model approach may lead to an "optimal" solution more often than the other approach. Of course, this puts greater pressure on the analyst to understand the causal mechanisms underlying the battlefield. The extent to which the analyst can do so will determine whether he succeeds or fails.

Order of Battle Analysis

Order of battle analysis involves putting together the enemy order of battle, its disposition, composition, strength, etc. The order of battle analyst relies heavily on enemy doctrine and information on composition of enemy units to put together the order of battle. For example, there are extensive templates that describe the doctrinal composition of units such as motorized rifle divisions and how they array themselves (their disposition) in battle. These templates tend to be generic for the different types of units at different echelons (e.g., tank regiments, motorized rifle battalions, 122 mm artillery battalions), with the details of specific units filled in through collection and analysis.

We found that OB analysts use this templating process extensively in their analysis. Typically, the OB analyst will receive a report concerning a type of unit (e.g., "Elements of the 3d Tank Battalion of the 122d Tank Regiment located at NB 315635"). They then look at their OB workbook to see what superordinate units this battalion belongs to. In addition, they may use the general template of how tank regiments array themselves in order to predict where other units of the regiment might be. This type of analysis might occur at an even lower level of analysis. For example, the analyst may receive a report regarding BTR tanks. He would then consult his order of battle workbook to see what units had such tanks and template from there.

Hence, the OB process is largely template-driven. In fact, OB analysts are often described as "walking encyclopedias" by their co-workers because of their wealth of content knowledge. Such knowledge extends not only to packages of concepts (e.g., frames such as motorized rifle divisions), but also to the individual

concepts that make up these packages (e.g., weapon systems, their ranges, lethality). Such "semantic" knowledge is important since it forms the building blocks of the templates the OB analysts use.

These templates are often used to develop inference rules. For example, if template A has components x, y, z and template B has components w, x, y, the analyst may develop an inference rules such as "If I see 'x', then it suggests either A or B." or "If I see 'x', then I should look for 'z' or 'w' but not 'y'." Such inferences rules give rise to two meta-rules which our OB analyst subjects seemed to use. First, there was a general rule "The more features of a template I see, the more likely that template is to be the case". The second rule related to discriminations between templates, "If my indicators (template features) appear to confirm more than one template, then look for indicators that are common to one but not the others".

OB analysts appeared to be aware that the world might not always conform to their template-based expectations. As a result, they often qualified their assessments with statements such as "It depends of the terrain" or "It depends on the mission". For example, an OB analyst might get a report concerning the location of an artillery unit. However, rather than plot the unit at the exact location specified by the report, the OB analyst might look at the map and, knowing that artillery units are typically placed so that they back up to woodlines for cover and concealment, plot the location of the unit by a woodline. This type of situation knowledge may integrate with the template knowledge in several ways. First, the analyst may have other templates or rules that specify how terrain and other factors influence OB (e.g., artillery is placed along woodlines). Here, the analyst would integrate the two templates (e.g., place the unit in this area but by the woods). The analyst may also have a more goal-means way of integrating the two. For example, the analyst may reason that the enemy wants concealment for his artillery. He may then reason as to what means friendly forces have for detecting his artillery and then look for means to conceal the artillery from the friendly detection means. This type of integration of templates and situations may involve a mental model analysis using goals and situation features (e.g., semantic knowledge).

Despite this inherent understanding that templates may not accurately describe all real world situations, OB analysts, nevertheless, rely heavily on such templates, which are based largely on enemy doctrine. While, we have no statistics (because probably none exist) on the frequency with which the enemy fights according to these doctrinal templates, it is useful to note that one of our subjects reported that he was once on an exercise where the opposing forces (OPFOR) deliberately task organized his force in a way that was totally inconsistent with doctrine. As a result, the blue forces (BLUEFOR) was totally confused as to who

they were fighting. The subject said it took a day and a half to finally figure out what the OPFOR was doing. The subject mentioned that had it been a real battle, a state of such confusion lasting a day and a half would have been devastating. We believe, then, that the templating approach, if used too heavily, could result in problems in novel situations.

Situation Assessment

Situation assessment is an extension of order of battle analysis where the goal is to take the OB information amassed by the OB analysts and use that information to predict enemy intentions and likely courses of action. This is primarily done (when done right) in support of friendly courses of action and friendly missions.

Situation assessment was the subject area most studied on the proposed project, and perhaps the most interesting. For one thing, unlike the other areas, the goal of situation assessment is to produce one answer: what will the enemy do? In OB analysis, there are many units being described; in target development, many targets are selected; in collection management, many assets are used to try to collect many pieces of information. As a result, these three domains provided more opportunities for "partial credit" and would seem to offer more opportunities for individual differences in terms of their solutions.

Situation assessment was an interesting domain in that there were as many differences as there were commonalties. While all subjects look at terrain and disposition of enemy forces (among other things), there appeared to be as many ways of interpreting such information as there were subjects. In addition, subjects (most of whom were reasonably experienced) came up with dramatically different conclusions as to what the enemy's most likely course of action was, even though they worked on the same problem using the same data. These differences ranged from being subtle (e.g., how would the enemy attack in a particular sector), to being more significant (e.g., would the enemy be most likely to attack in the north or the south), to being rather fundamental (e.g., is the enemy going to attack or defend).

One interesting observation is worth making about the situation assessment process (which may also explain the dramatic differences in situation assessment processes and conclusions). While, situation assessment is technically an inference process (i.e., inferring conclusions from the situational data), many subjects appeared to approach situation assessment as a planning process (i.e., If I were the Red commander, what course of action would I develop in this situation?). As a result, subjects may have come up with different plans and then attributed these to the enemy. Currently, DSC is pursuing a separate line of research

looking at the G-3 planning process. In an ongoing study, we have run sixteen subjects and asked them to recommend blue courses of action in a tactical scenario. There have been nearly as many different recommended courses of action as there are subjects! This seems to support our contention that subjects doing situation assessment are acting more like planners than "inferencers."

More support for this contention is gathered by examining some of the situation assessment strategies used by the subjects. Virtually, all subjects begin with a general background assessment that includes reviewing the OB products, e.g., enemy disposition, composition, and strength. Subjects also review the terrain to see what avenues of approach exist between friendly and enemy objectives. Here, our subjects start to diverge.

Perhaps our two best and most experienced subjects (one was a division G-2--the only one we interviewed, the other was warrant officer with some 22 years of experience) used a "wargaming" approach to situation assessment. The tried to envision likely blue and red plans and pit the two against each other. These subjects integrated friendly and perceived enemy missions into their analysis. As a result, they developed likely red courses of action that but would meet perceived enemy missions and respond to likely blue courses of action.

The wargaming approach was highly dynamic in nature. The two subjects appeared to evolve plans as they analyzed the problem as opposed to applying "canned" plans to the situation (although they may have had a set of standard considerations they brought to bear like dealing with reserve forces). This approach appeared largely goal-based, perhaps using mental models as a means for generating action sequences to support those goals. Unfortunately, one subject was working on an offensive scenario while the other was working on a defensive scenario so that the two "results" cannot be compared.

A second type of strategy appeared to be best characterized as a "bean-counting" approach. Here, subjects used notions of combat power ratios and "massing for the attack" to project likely enemy course of action. Here, subjects would focus on relative strengths of enemy forces, looking at areas where the enemy had massed or narrowed its sector to increase combat power relative to opposing forces. In such cases, subjects would predict enemy main attack in places where relative combat power was greatest and the enemy had massed for the attack. Unlike the wargaming approach, subjects in these cases did not develop the enemy course of action in great detail, nor were explicit blue plans and friendly and enemy contingencies considered.

The "bean-counting" approach appeared to be largely production rule-based. Here, subjects would go through computational formulas (although some did crude passes at them rather than

perform the extensive analysis) to generate assessments of relative combat powers. In our main report, we discuss how this approach was so explicitly procedural that the formula could be captured and given to a novice who in turn could emulate the subject.

A third strategy could be viewed as doctrine-based templating. Here, subjects would rely heavily on doctrinal considerations such as "the enemy will reinforce success" to predict location of enemy main attack. In addition, subjects looked heavily for specific indicators of course of action such as: positioning of reserve; depth of artillery, air defense, command posts; location and type of engineering assets, etc. In particular, disposition of forces played a major role, e.g., was the enemy postured for attack or defense. These assessments were largely template-driven, based on Soviet doctrine.

These three strategies appeared to correlate roughly with the experience of the subjects involved. As discussed above, the two best subjects used the wargaming strategy in situation assessment. The least experienced or knowledgeable subjects appeared to rely heavily on the template/indicator method. It should be noted however, that these strategies are not totally orthogonal. For example, all subjects paid some attention to indicators or combat power. However, more experienced subjects appeared to build on what the less experienced subjects did.

Perhaps the template/indicator situation assessment strategy was the most simplistic. While undoubtedly most analysts feel that Soviet doctrine is a good basis for situation assessment (subjects agree that Soviets train very much according to their doctrine), they agree that deception is an integral part of their warfighting. Therefore, an overreliance on templates or doctrine-based indicators may be highly susceptible to deception efforts.

The "bean-counting" approach may be a more powerful situation assessment strategy since it focuses more deeply on what the Soviet's capabilities are. While this approach may not guarantee a correct inference about what the enemy <u>will</u> do, it may help blue prepare for what they <u>can</u> do. Still, such a strategy may not be optimal in helping blue prepare for <u>how</u> the enemy will fight, even in cases where the analyst has successfully predicted where and when the enemy will fight.

The mental model approach appeared to be the most sophisticated of the strategies used. While this approach appeared to rely less on "hard data," it appeared to utilize more of what the subjects knew about warfighting. Importantly, this strategy got into the "whys and wherefores" of the battle in order to attribute purpose to the enemy's actions. In doing so, it provided a more detailed analysis of what the enemy was expected to do and what blue could do in turn. Our understanding from

discussions with our subjects is that this is what the G-2 is supposed to do if he is doing his job right.

The mental model approach is subject to biases as well. The literature is replete with studies regarding planning and inference biases. For example, people often suppress uncertainty (cf., Kahneman, Slovic, & Tversky, 1982). Tolcott et al. (1987) found that subjects are often slow to update their assessment of a situation, even when new information conflicts with their old impressions. Leddo and Govedich (1986) found that planners often ignore external events that can impact plans (the G-2's task is to prevent this from happening, but if he is acting as a planner, he might be susceptible to the same bias), rarely question the validity of their assumptions, and generate plans with little "branching" (i.e., generate plans with a single thread that provides little in the way of midcourse corrections). While, we have not established that our subjects exhibited such biases, if they did, it might be most likely that they would occur by those subjects using the mental model/goal-based approach since their strategy most closely emulates planning. In such cases, subjects who engage in a more "bean-counting," or indicator-based approach, i.e., they are more data-driven, more be less susceptible to such biases.

Summary of Findings

The first matrix below presents a summary of the different domains, different problem solving strategies exhibited by subjects in each domain, and conditions that seemed to influence which strategy subjects used.

Functional Area	Strategies Observed	Factors Assoc. w. Strategies
Collection Management	Frame	Less experienced
	Goal/Plan	More experienced
Target Development	Semantic/Prod Rule	Unit (Fort Hood) or theater. (Europe-doctrine well known).
	Mental Model	Unit (Fort Lewis) or theater. (Korea-doctrine less known).
Order of Battle Anal	Frame/Prod Rule	General strategy
Situation Assessment	Goal/Mental Model	Most experienced
	Production Rule	Middle experienced
	Frame	Least experienced

The second matrix below presents a summary of the strengths and weaknesses associated with using each individual strategy. The extent to which subjects used hybrid strategies (e.g., semantic knowledge and production rules) will determine the combination of strengths and weaknesses of that subject's approach.

Strategy	Strengths	Weaknesses
Frames	Balanced, less subject to extremes, biases. Compiled and efficient (may be good under time stress).	Little attention to situation specifics. Relatively inflexible.
Semantic	Compiled and efficient (may be good under time stress). Balanced, less subject to extremes and biases.	Little attention to situation specifics. Relatively inflexible.
Production Rules	Replicable. Produces concrete output.	Tend to miss "big, picture," intent.
Goal/Plan	Focuses on situation needs.	Susceptible to biases.
Mental Model	Focuses on situation needs and may produce optimal solution.	Requires getting model right (may get on wrong track). Cognitively intensive and inefficient.

Discussion

Our findings, particularly those in the situation assessment domain, may shed some light on how expertise is acquired. Anderson (1982) discusses how people develop certain kinds of skill knowledge. He makes a distinction between declarative and procedural knowledge. Declarative knowledge is largely factual knowledge and is typically represented semantically. Declarative knowledge has been described as "knowing that." Procedural knowledge is "knowing how," i.e., the steps necessary in order to perform some function. Procedural knowledge represents the operationalization of declarative knowledge where knowledge is transformed from an abstract, factual representation to a more concrete, procedural representation. Anderson uses production rules to represent procedural knowledge.

We feel our findings may extend Anderson's framework. We also found that relatively inexperienced subjects used templating or semantic (frame) approaches to performing their jobs. However, more experienced subjects (in the case of the situation assessment subjects) used a more procedural or production rule approach. However, our most experienced subjects, those with more than twenty years of experience and who had achieved a high level of domain expertise) exhibited a more goal-based and mental model (causal analysis) approach. We can characterize this approach as "knowing why."

We can then say that expertise evolves first through "knowing that" (i.e., what one should do), to "knowing how" (i.e., how to do it), to "knowing why." If this is the case, it suggests that people acquire knowledge at a general level (the declarative knowledge) at first. This knowledge then grows more specific, concrete, and proceduralized as people learn how to apply that knowledge to specific problems. Finally, the knowledge develops to an even more abstract level than it was originally as people learn the underlying causal mechanisms and the goals involved in their domain. These high-level goals and mental models then dominate the processing in experts.

This theory of knowledge development is compatible with Cohen's (1989) framework. Cohen uses Rosch's (1983) concept of "basic level concept" to argue that novices start out with basic level concepts. This knowledge then evolves in two directions, up (toward greater levels of abstraction) and down (toward greater levels of detail). The present findings support Cohen's notion with the qualification that the process of proceduralizing (concertizing) knowledge appears to happen either prior to or quicker than the process of abstracting knowledge.

These findings suggest some interesting implications for training. To the best of our knowledge, people become military experts through experience, not through training. Courses typically provide background rather than extensive training to build expertise. A soldier may spend only a few years of his career in school. Most of that training is focused on acquiring declarative knowledge (the student learns doctrines and standard procedures). Exercises subsequently focus on building procedural knowledge, i.e., how to use the classroom knowledge in realistic situations. Little emphasis is given on the abstraction of knowledge, or that knowledge which distinguishes experts from non-experts. Hence, such personnel must try to develop expertise on his own. However, we argue that the process of acquiring expertise could be facilitated in classroom training by emphasizing both the proceduralization and abstraction of declarative knowledge, with special emphasis on the abstraction.

General Discussion

The goals of the proposed project were twofold. First, we wanted to gain insight into how experts represented and used knowledge. Second, we wanted to develop a set of knowledge elicitation techniques that could be used integratively and were driven by an understanding of how experts represent and use knowledge.

Our research suggests that experts possess a diverse range of knowledge, as illustrated by the various knowledge structures that comprise our INKS framework. This knowledge is used integratively and experts often flow from one structure to another or even transform knowledge from one structure to another in order to accomplish a specific task.

While it may not be counterintuitive that experts possess a wide range of knowledge structures or that they are used integratively, we feel that our findings underscore the importance of conceptualizing expert knowledge as entailing integrated knowledge rather than single structures. While there is much research individually supporting the diverse structures used in our INKS framework, most psychologists and AI practitioners focus on single representation schemes in psychological theories and AI systems. In other words, while it should be well known that experts possess a variety of knowledge structures, in practice, experts are treated as if they possess only one type of knowledge. The most notable exception to this is John Anderson's ACT* theory which explores the transition of semantic knowledge (declarative knowledge) to production rule knowledge (procedural knowledge). Our INKS framework is consistent with this but extends the complexity of the knowledge used even further.

We hope that our work in this project will provide an added push to the psychology and AI communities so that they will utilize integrated knowledge representation schemes even more. We argue that this will be of practical benefit as well. Two important applications served by understanding how experts represent knowledge are training and the development of expert and decision support systems.

Perhaps the most important application served by expert knowledge models is developing methods for training people to become experts. The Army invests huge sums of money in training its personnel. Currently, people acquire expertise after long years of experience. Given that many people do not maintain lengthy Army careers and those that do may only be considered experts for a few short years before they retire, it is well worth the effort to develop methods for helping personnel to acquire expertise more quickly and to achieve higher levels of competence.

We believe our INKS framework can serve as the basis for instructional methods. In particular, an INKS-based model of expert knowledge can serve as a target for instruction. The INKS model defines a set of knowledge and problem-solving skills that the student is trying to emulate. In Leddo, Sak, and Laskey (1989), we use this INKS framework and our knowledge elicitation framework as a basis for building an Intelligent Tutoring System called THINKER.

The goal behind THINKER is to teach students to represent and use knowledge as experts do. Hence, THINKER's curriculum is based on a model of how experts represent and use knowledge. This model was developed by conducting knowledge elicitation on subject matter experts in the domain THINKER uses: Follow-on Force Attack (FOFA), which is the Air Force's analog to the Army's target development). The elicited knowledge was then placed into an INKS representation.

THINKER uses a three step instructional procedure to build integrated problem solving in students. The philosophy behind this procedure is that problem solving requires two major steps. The first step is analyzing the problem to infer the relevant problem features and how they relate to the problem objective. This is then used to select a problem-solving strategy. The second step is applying the problem-solving strategy to the particular problem in order to generate a solution.

We regard the first step as a "bottom-up" process, linking low-level problem features to high-level problem-solving strategies. The second step can be regarded as a "top-down" process, where high-level problem-solving strategies are mapped onto concrete situations.

As a result, the first instructional procedure that THINKER uses is to provide students with an overview of the domain. This establishes a cognitive framework for the material that the students will learn in the remaining lessons. In the overview, students learn about the objectives, plans and causal reasoning associated with the domains. These get fleshed out in later lessons.

The second instructional procedure is called "bottom-up" instruction. Here, the students learn the fundamental concepts that comprise the domain and how they relate to specific planning strategies. In THINKER, these concepts relate to the terrain the enemy uses and how it affects the enemy's ability to conduct movement to impede friendly objectives. In this stage of instruction, students to not yet implement plans.

The third instructional procedure is called "top-down" instruction. Here, students learn how to apply different plans in specific problem situations. In THINKER, this translates to developing and executing targeting plans.

The key behind THINKER's instructional approach is the interplay between high-level planning knowledge and low-level feature knowledge. This is strengthened by teaching students the causal relationships between the plans they choose and how they lead to goal-relevant outcomes. However, THINKER demonstrates a new application for the INKS framework and the resulting knowledge elicitation techniques, namely, the evaluation of training.

If the goal of instruction is to teach students to emulate experts, then not only is it important to assess performance, but also one must assess whether students are representing and using knowledge as experts do. THINKER evaluates training by using a simplified version of the Cognitive Structure Analysis technique to probe students to see how they represent and use knowledge. The students' knowledge is then compared to the expert knowledge model to determine what discrepancies exist. This comparison of students' knowledge to expert knowledge allows for the presentation of remedial instruction that is tailored to the specific knowledge requirements of each student.

We believe THINKER represents a good first step toward using results such as those obtained in the present project to apply theories of integrated knowledge representation and knowledge elicitation techniques to the training domain.

A second application of the INKS framework lies in the area of developing expert systems. Currently, there are two main categories of expert systems: rule-based and frame-based. These are primarily driven by considerations of system architecture. There have been some recent attempts at developing systems that are both rule-based and frame-based, but use of these structures seems to be driven by considerations of computational efficiency rather than psychological validity.

We argue that expert systems could be made more powerful and more compatible by using INKS-like representations that more closely match how experts represent and use knowledge. Such systems could handle a more diverse range of knowledge inputs and would undoubtedly be capable of handling a diverse range of reasoning tasks.

The power of an INKS framework is demonstrated in a Decision Science Consortium system called INKS (Leddo, Cardie, & Abelson, 1987). INKS reads stories that are parsed into Schank's (1972) Conceptual Dependency primitives. INKS diagnoses planning failures (i.e., the actor fails to achieve his goal), suggests potential solutions to the failure, and updates its memory when successful solutions occur. These successful solutions are applied when similar failures occur in other stories.

INKS demonstrates the power of an integrated knowledge representation scheme in that it uses MOPs to process the stories, mental models and semantic knowledge to generate explanations as to why the failure occurred and to generate candidate solutions (which are typically in the form of production rules), and updates its memory by inserting new production rules or scenes into the MOPs or changing semantic knowledge to deal with future planning failures. INKS exhibits understanding, explanation, planning, and learning all by using its integrated knowledge structure scheme. It is rare for AI systems to exhibit all of these capabilities in one system.

Our second goal in the current project was to develop a set of knowledge elicitation techniques that would complement each other and be designed to produce psychologically valid models of expert knowledge. Previous work in knowledge elicitation has been driven more by the application that knowledge would serve (e.g., expert system) than by considerations of how the expert represents and uses knowledge. In the past, knowledge engineers would typically predefine the representation framework to be used as a model of "expert" knowledge, and force fit the elicited knowledge (which was often taken in a think-aloud protocol) into that framework.

Our approach was to develop a set of elicitation techniques that could elicit different types of knowledge that form the basis of expert problem solving and then develop a means for assessing how that knowledge was structured in the expert's head. Our integrated knowledge elicitation methodology was designed to accomplish these goals. We refined the traditional think-aloud protocol technique to increase its power. This was done by developing specific variations to focus experts on different aspects of knowledge (i.e., group vs individual, full-scale vs mini problems, context- vs data-driven). As was done in previous usages of the think-aloud technique, our usage was designed to elicit general problem solving strategies and procedures.

This standard problem solving technique was then supplemented with additional techniques that are more specialized in the knowledge they elicit. The Project Ahead technique elicited the scenario-generation knowledge subjects have as they simulate how problems unfold. This in turn guides decisions that the subjects make. The Conflict Resolution technique elicits how the experts reason with situational information, generate conclusions and resolve uncertainty. The Project Ahead and Conflict Resolution techniques, then, are designed to elicit more deeply how experts solve problems than might be elicited through a think-aloud technique alone.

Finally, Cognitive Structure Analysis is used to infer the underlying representations and relationships behind the knowledge the experts use. We believe CSA represents a significant

conceptual as well as technological innovation in the knowledge elicitation field. Without the CSA technique, our use of the other techniques may leave us in a similar situation to other knowledge engineers, who have elicited a set of knowledge but still have limited criteria by which to represent that knowledge. In that respect, we might be accused of force-fitting our elicited knowledge into our INKS scheme. CSA attempts to validate the underlying structure of the elicited knowledge.

We acknowledge that more work is needed on the CSA technique. There are undoubtedly a wider range of procedures that we can use to validate the knowledge structures used by experts. Future variations of CSA may be more rigorous in that they may employ cognitive psychology paradigms involving priming and recall to infer underlying structural relationships of knowledge.

On the other hand, the CSA technique is a good first pass at validation of knowledge structures elicited from experts. We hope that CSA will induce future generations of knowledge engineers to validate their knowledge models as well. Why validate knowledge models at all? The key premise behind our elicitation approach is to uncover not only what experts do, but how and why they do it. We feel that this is the key to generate models of expertise that extend beyond the test cases under which the "what" knowledge is elicited. We feel it is key to replicate the underlying process as well as the overt behavior. Otherwise, when new situations are encountered, the expert model will collapse. This is perhaps the best explanation as to why expert systems are typically brittle and apply to only a limited set of cases.

We argue that developing valid models of expert knowledge and using them to drive expert and decision support systems will increase their power as well as their compatibility with experts. Ideally, the knowledge elicitors will serve as the driving force behind a new generation of system designs.

Another area where knowledge elicitation and knowledge modeling can drive a field is in training and training evaluation. Currently, training is designed to produce a target level of performance, i.e., it is output oriented. There is little focus on teaching people to replicate the thought processes of experts. However, current educational thinking (see Costa, 1985, for a review) advocates focusing on problem solving process as well as output.

We feel the approach to training outlined in the THINKER tutor exemplifies this knowledge-based approach which utilizes knowledge elicitation technologies. Recall that THINKER's domain knowledge is developed from elicitation sessions with experts rather than from a textbook. We feel this approach has several benefits. First, the organization of the material has already been performed

by the expert. Textbooks are primarily sequentially organized, leaving the student to organize the material in his head. Expert knowledge is primarily hierarchically organized with the relevant interconnections between concepts and procedures already established. Second, textbooks emphasize breadth. Subject areas are covered very broadly. It is clear that students are unable to memorize everything they read, yet it is difficult for them to decide which material to emphasize and which material they can discard. The expert has already gone through the process of sifting out the relevant knowledge and discarding the less relevant knowledge. Third, the expert is likely to have acquired a great deal of practical knowledge that is often missing in textbooks. This knowledge may be captured during the knowledge elicitation sessions, but left out of the books.

THINKER also used knowledge elicitation techniques to evaluate students' learning. This was done in addition to traditional performance measures, e.g., how many answers the student got right or wrong. We argue that this type of assessment is valuable for several reasons. First, students can generate answers for a variety of different reasons. Hence, a student can apply a formula incorrectly and still generate the right answer. It is true that we would not expect this pattern to generalize, i.e., a student consistently guessing right. On the other hand, the student may get the wrong answer even though he essentially understood the right formula, but made a careless error along the way. Hence, the performance method of evaluation does not distinguish correct guesses from correct reasoning and careless mistakes from incorrect reasoning. A knowledge elicitation approach solves this problem.

Second, many domains do not necessarily have readily objective answers by which performance can be evaluated. In the military domain, most problem solving is hypothetical in nature. While there may be school solutions, there is no guarantee that those solutions are both effective and the only possible solutions in an actual wartime environment. In such cases, evaluating the reasoning strategy rather than the actual performance may be more diagnostic. Our own research (noted above) suggests that experienced Army problem solvers often come up with dramatically different solutions to the same problem. It is hard to imagine that no more than one of them is right.

Third, simply evaluating performance does little to suggest ways to improve performance. Knowing that someone has erred suggests that the person needs remedial instruction. However, unless the causes for the error are known, what remedial instruction to give remains unclear. As THINKER demonstrates, knowledge elicitation techniques can be used to identify where the gaps in knowledge or understanding lie. This, in turn, can be used to drive how remedial instruction is delivered, tailoring that instruction to the specific needs of the students.

An additional benefit of the proposed project was some insights into what makes someone an expert and how expertise may be acquired. By comparing the problem solving strategies of subjects at different skill and experience levels, we may see how problem solving strategies evolve over time. Unfortunately, our test is not rigorous as the scope of the present project did not allow a longitudinal study involving skill development within subject. However, our findings may serve as the basis for future research.

Based on our findings, we believe that people who are true experts in their fields use goals and causal mental models as their primary means for solving problems. Goals or objectives appear to drive the problem solving approach and the types of knowledge they use. The first part of their strategy is to define their goals and to assess the situation relevant to those goals. This assessment is then used to generate a plan to achieve the desired objectives, i.e., influence the situation so that it transforms into a state that is consistent with the goal state. This planning process appears to be causally-driven. The expert understands the relationships between the action he takes and the outcomes that will be produced. In essence the expert performs a simulation in his mind. This simulation may be an iterative process as the expert refines his plan in order to optimize the outcome.

We characterize this process as knowing "why." Unlike Anderson's concept of knowing "how," experts appear to act deliberately, rather than mechanically. The expert appears to innovate, rather than apply fixed plans to a given situation. This innovation may not be a fixed trait of experts, but rather a by-product of the fact that no two problems are identical and the expert uses his causal knowledge and his goals to produce an optimal solution to the specific problem.

We contrast this style of problem solving with the strategies used by problem solvers of less experience (e.g., novices). At the lowest level of skill and experience, subjects appeared to be primarily template-driven. They applied fixed patterns of inference and planning to problems they encountered. Subjects who were more experienced, but perhaps not experts, seemed to be more procedural. They took features of the situation and tried to draw specific conclusions from them and responded to those conclusions (e.g., the situation assessment subjects who counted units and strengths). While this approach was more tailored to the specific problems at hand than the template approach, it lacked the "top-down," goal-driven force that the most experienced and skilled subjects used.

It is difficult to say whether these strategies evolve in sequence. Certainly, not everyone reaches the level of expertise that are best subjects achieved. In our sample, the number of

subjects who exhibited that goal/causal strategy was rather low. We do not know if they evolved this strategy by going through the other strategies or whether they always had the "seeds of greatness." However, the logical sequence of templates to procedures to goals and causal reasoning carries intuitive appeal.

It seems reasonable that a young intelligence officer or enlisted person would go through the standard schooling. Here, he learns about Soviet doctrine, standard operating procedures (SOPs), and principles of warfighting. It makes sense that in his early experience, he would then look to see if he recognized any of these principles or templates in the problems he encountered. At this point the soldier has little experience so he probably will do little more than apply what he has learned in the classroom, i.e., the generic, doctrinal solutions.

As he gains experience, the soldier will undoubtedly learn that different problem situations have nuances in them that should be responded to with different actions. Here, the soldier learns new procedures and even how to generate procedures to deal with these situations. He retains his templates, but now can adapt them more dynamically to specific situations. In fact the templates can serve as a basis for generating procedures as was illustrated by the order of battle analysts.

At some point, the soldier may realize that there are driving forces behind the situations and actions that he observes. He realizes that the templates are driven by objectives (both friendly and enemy) and that the plans be works with operate in a sophisticated environment. Here, the soldier begins learning to think in terms of goals and objectives and how the causal environment he operates in affects these.

This is undoubtedly a lengthy learning process. The goals are often complex as they are numerous, are possessed by different players, have different levels of importance and may compete or help each other. Here, the soldier must learn to recognize the important goals and be able to understand how his own goals relate to enemy goals. We see evidence of these considerations in our subjects. Some subjects think primarily of division-level objectives (when given a division-level problem). Others try to fit their division-level objectives into the Corps perspective and the Corps goals. Our most advanced subjects also fit friendly objectives with enemy objectives and consider the adversarial nature of both and how that impacts on friendly plans.

Once an understanding of goals has been achieved, the soldier must learn how the plans and courses of action he adopts will impact these goals. Not only does the soldier need causal planning knowledge, but he must also understand his goals thoroughly. All subjects realize that they must support the

Commander's mission, but only the most experienced and skills really give deep thought to what the mission means and what the true objectives are. In fact, this may lead the subject to interpret the mission (goals) in ways that most subjects would not pick up on.

The resulting causal understanding of the domain that the soldier would develop would necessarily be driven by the understanding of goals. Both friendly and enemy plans are understood in terms of these goals. Any impact of events would be processed in terms of how they affect the attainment of the goals. Undoubtedly, the understanding of goals would focus the soldier's attention on specific questions about the domain, which in turn would shape the knowledge he would acquire and how it would be organized in his head.

Over time, then, the soldier's knowledge would be organized in a highly "top-down" fashion, centered around his goals. Previous knowledge such as templates and procedures would now be driven by a goal-based organization. At this point, we would argue that the soldier had acquired expertise.

It is our hope that these findings, collectively, could be used to facilitate the process by which people acquire expertise. We feel that our suggestions regarding training show promise and would lead to innovations in the field of training. We also hope that our approaches to knowledge representation and use, particularly the INKS framework, will encourage new approaches to the development of expert and AI systems that are more powerful and more compatible with its human users. Finally, we hope that the issues we have raised here will stimulate a wave of theoretical and empirical work that is designed to get at some of the answers to our questions.

References

Adelson, B. (1984). When novices surpass experts: The difficulty of a task may increase with expertise. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10(3), 484-495.

Anderson, J. R. (1980). Cognitive psychology and its implications. San Francisco, CA: W.H. Freeman and Co.

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review, 89, 369-406.

Berry, D. C. & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalisable knowledge. Quarterly Journal of Experimental Psychology, 36A, pp. 209-231.

Brown, R. V. (1989). Normative models for eliciting and exploiting intelligence knowledge (Technical Report 89-4). Reston, VA: Decision Science Consortium, Inc.

Chase, W. G. & Simon, H. A. (1973). The mind's eye in chess. In W.G. Chase (Ed.), Visual information processing. NY: Academic Press.

Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121-152.

Cohen, M. S. (1989). A cognitive theory of knowledge elicitation (Working Paper). Reston, VA: Decision Science Consortium, Inc.

Costa, A. L. (Ed.) (1985). Developing minds: A resource book for teaching thinking. Alexandria, VA: Association for Supervision and Curriculum Development.

DeKleer, J. & Brown, J. S. (1981). Mental models of physical mechanisms and their acquisition. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. Psychological Review, 86, 465-485.

Ericsson, K. A. & Simon, H. A. (1988). Verbal reports as data. Psychological Review, 87(3), 215-251.

Ericsson, K. A. & Simon, H. A. (1984). Protocol analysis: Verbal reports as data. Cambridge, MA: The MIT Press.

Galambos, J. A. (1986). Knowledge structures for common activities. In Galambos, J. A., Abelson, R. P., and Black, J. B. (Eds.) Knowledge structures. Hillsdale, NJ: Erlbaum.

Greeno, J. G. (1983). Conceptual entities. In Gentner, D. and Stevens, A. L. (Eds.), Mental Models, pp. 227-252. Hillsdale, NJ: Lawrence Erlbaum Assoc.

Hayes-Roth, F., Waterman, D. A., and Lenat, D. B. (1983). Building expert systems. Reading: MA: Addison-Wesley Publishing Co., Inc.

Johnson-Laird, P. N. (1983). Mental models. Cambridge, MA: Harvard University Press.

Kadane, J. B. & Lichtenstein, S. (1982). A subjectivist view of calibration (Report 82-6). Eugene, OR: Decision Research.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). Judgment under uncertainty: Heuristics and biases. NY: Cambridge University Press.

Larkin, J. L. (1981). Enriching formal knowledge: A model for learning to solve textbook physics problems. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Larkin, J. L., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. Science,208, 1335-1342.

Leddo, J. M. & Cohen, M. S. (1987). A cognitive science approach to elicitation of expert knowledge. Proceedings of the 1987 JDL Command and Control Research Symposium. McLean, VA: Science Applications International Corporation.

Leddo, J. & Cohen, M. S. (1989). Cognitive structure analysis: A technique for eliciting the content and structure of expert knowledge. Proceedings of 1989 AI Systems in Government Conference. McLean, VA: The MITRE Corporation.

Leddo, J. & Govedich M. M. (1986). Role specific cognitive approaches to decision making. Falls Church, VA: Decision Science Consortium.

Leddo, J. M., Mullin, T. M., Cohen, M. S., Bresnick, T. A., Marvin, F. F., & O'Connor, M. F. (1988). Knowledge Elicitation: Phase I Final Report, Vols I & II (Technical Report 87-15). Reston, VA: Decision Science Consortium, Inc.

Leddo, J. M., Cardie, C. T., & Abelson, R. P. (1987). An integrated framework for knowledge representation and acquisition (Technical Report 87-14). Falls Church, VA: Decision Science Consortium, Inc.

Leddo, J., Cohen, M. S., Bresnick, T. A., Marvin, F. F. (1989). Eliciting expert knowledge using problem solving techniques (Working Paper). Reston, VA: Decision Science Consortium, Inc.

Leddo, J. M., Sak, S. G., & Laskey, K. B. (1989). THINKER: An intelligent tutoring system for building expert knowledge (Technical Report 89-7). Reston, VA: Decision Science Consortium, Inc.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. NY: Cambridge University Press.

Martin, A. W., Mullin, T. M., & Leddo, J. M. (1989). Learning from experienced analysts (LFEA) phase zero: Results of working conference on expertise in intelligence analysis. Reston, VA: Decision Science Consortium, Inc.

McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. Cognitive Psychology, 13.

Minsky, M. (1975). A framework for representing knowledge. In P. Winston (Ed.), The psychology of computer vision. NY: McGraw-Hill.

Newell, A. Rosenbloom, P. S. Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), Cognitive skills and their acquisition, pp. 1-55. Hillsdale, NJ: Erlbaum Assoc.

Newell, A. & Simon, H. A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice-Hall.

Novack, G. S. & Araya, A. A. (1981). Physics problem solving using multiple views (TR 173). Austin, TX: University of Texas at Austin.

Quillian, M.R. (1966). Semantic memory. Cambridge, MA: Bolt, Beranek and Newman.

Rosch, E. (1983). Prototype classification and logical classification: The two systems. In E. Scholnick (Ed.), New trends in cognitive representation: Challenges to Piaget's theory. Hillsdale, NJ: Lawrence Erlbaum Associates

Schank, R. C. (1972). Conceptual dependency: A theory of natural language understanding. Cognitive Psychology, 3, 552 -631.

Schank, R.C. (1982). Dynamic memory: A theory of reminding and learning in computers and people. MA: Cambridge University Press.

Schank, R. C. & Abelson, R. P. (1977). Scripts, plans, goals and understanding: An inquiry into human knowledge structures. Hillsdale, NJ: Erlbaum.

Schum, D. A. (1980). Current developments in research on cascaded inference processes. In T. S. Wallsten (Ed.), Cognitive process in choice and decision behavior. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shoenfeld, A. H. & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8(5), 484-494.

Tolcott, M. A., Marvin, F. F., & Lehner, P. E. (1987). Effects of early decisions on later judgments in an evolving situation (Technical Report 87-10). Falls Church, VA: Decision Science Consortium, Inc.

Tolcott, M. A., Marvin, F. F., & Bresnick, T. A. (1981). The confirmation bias in evolving decisions. Proceedings of 1989 Symposium on Command and Control Research.

Tversky, A. & Kahneman, D. (1083). Extensional vs. intuitive reasoning: The conjunctive fallacy in probability judgment. Psychological Review, 90, 193-315.

Waterman, D. A. & Newell, A. (1971). Protocol analysis as a task for artificial intelligence. Artificial Intelligence, 285 -318.

Weiser, M. & Shertz, J. (1983). Programming problem representation in novice and expert programmers. International Journal of Man-Machine Studies, 19.

Wilensky, R. (1983). Planning and understanding: A computational approach to human reasoning. Reading, MA: Addison-Welsey.

Winograd, T. (1975). Frame representations and the declarative/procedural controversy. In D. Bobrow & A. Collins (Eds.), Representation and understanding: Studies in cognitive science. New York: Academic Press.

APPENDIX A

Procedures for Carrying out Knowledge Elicitation Using an Integrated Knowledge Elicitation Methodology

Overview

In this appendix, we outline the four techniques that comprise the integrated knowledge elicitation methodology (Cognitive Structure Analysis, Problem Solving, Conflict Resolution, and Project Ahead), lay out guidelines for using this methodology in conducting knowledge elicitation, and give procedures for building models of the elicited knowledge.

Knowledge Elicitation Techniques

Cognitive Structure Analysis

Cognitive Structure Analysis (CSA) is a knowledge elicitation technique which utilizes an interview (question and answer) format. CSA is designed to identify and flesh out knowledge structures that experts use. The basic philosophy behind CSA is that different knowledge structures have different implications for how knowledge will be most naturally represented and used. Testing these implications allows the knowledge engineer to infer the identity of the underlying knowledge structures. Hence, there are two major steps to CSA: (1) discriminating among knowledge structures, and (2) filling in the content of the selected structures. We discuss each of these steps in turn.

Identification and Discrimination of Knowledge Structures. Suppose we wished to test whether an expert represents a particular sequence of events as either a MOP or as a production rule chain. MOPs are organized hierarchically while rules tend to have a flat organization. Knowledge in a hierarchical representation is accessed according its relationship to top-level goals while knowledge in a flat representation is accessed by serial search. The implications for these two modes of representations are that if an expert is organizing knowledge using a MOP, then "top-level" knowledge should be quickest and easiest to retrieve, regardless of its temporal location. If the expert is using production rules, then speed of access should depend upon sequence, not relative importance to the sequence.

For example, if knowledge of going to restaurants is represented using a MOP, then a person should be able to respond "yes" to the question, "Do you pay at restaurants?" more quickly than "Do you use forks at restaurants?" If rules are used, then the speed of response would be the opposite (assuming a forward

chaining search). Galambos and Rips (1982) have used such a methodology to demonstrate that many routine sequences of events (e.g., changing a flat tire) are organized by MOP-like representations.

Hence, the first step in Cognitive Structure Analysis is to try to anticipate the likely types of knowledge structures that experts may have and determine what factors would discriminate among them. While it is certainly the case that experts may have unforeseen ways of representing knowledge, we find that having a set of discriminators is a useful way to develop hypotheses about characteristics of new structures, even if we cannot specify what they are in advance. Hence, CSA can be used in an exploratory as well as confirmatory mode.

We selected knowledge structures to be discriminated based on findings in the psychology and cognitive science literature. Our criteria were that these structures should have received some empirical support and played complementary and somewhat distinct roles in representing knowledge. In all, five representation frameworks were adopted. The knowledge structures chosen were MOPs (event-based schema, Schank, 1982), production rules (which organize situation-specific procedures), semantic nets (which organize feature knowledge and categorical relationships among concepts), object frames (a sub-category of frames which focuses on collections of objects, e.g., kitchen, motorized rifle division), and mental models (mental models as we use them refer to structures which capture the causal knowledge that underlies the production of events, cf. Leddo & Cohen, 1987; Leddo, Cardie, & Abelson, 1987; Leddo, et al., 1988).

We view these five frameworks as partial and complementary approaches rather than as mutually exclusive alternatives for representing the same knowledge (although each has often been seen as sufficient by itself and a competitor to the others). MOPs (or more generally, event-based schema) capture high-level goal and planning knowledge. Production rules are most effective for capturing situation-specific procedural knowledge that might contribute to general planning knowledge. Semantic nets capture object/concept knowledge, while object frames capture knowledge regarding collections of objects. Mental models show causal relationships and how they produce events (traditionally, causal knowledge has been implicit at best in event schema and virtually non-existent in production rules). Mental models play an important role in that they can explain why given procedures are executed and also link event knowledge to content knowledge (the reasons for carrying out events often depend on properties of objects, e.g., we ask managers for help when we are having trouble getting service because they are powerful and have authority).

Once we have identified knowledge structures of interest, the next step is to develop a set of discriminators for them. We propose three key discriminators and show how they can be used for the knowledge structures listed above. These are the structures' basic units, organizing principle, and inferential use.

Basic units refers to whether the structures are organized around events or objects. We view MOPs and production rules as event oriented (since they typically capture procedural knowledge) and object frames and semantic nets as object oriented (since they typically describe states of the world). Object frames may be further discriminated from semantic nets in that object frames are viewed as organizing collections of objects rather than single objects. Mental models are often organized around objects, but emphasize how these objects interact to produce events. Hence, we argue that mental models are organized around objects <u>and</u> events. As such, this distinguishes MOPs and production rules from object frames and semantic nets, but is perhaps less useful in discriminating mental models from other structures.

Organizing principles refers to what drives the overall organization of the knowledge. As discussed earlier, MOPs organize goal-based knowledge. Hence, goals can serve as an organizing principle. We would argue that object frames are typically organized around some inherent goal or purpose. For example, kitchens are designed to fulfill a certain purpose, while motorized rifle divisions are organized to carry out specific missions.

Another organizing principle is causal relationships as typified by mental models. Here, goals are not necessary as one can have a mental model about such things as why objects fall to earth when dropped or how electricity works.

A third organizing principle refers to temporal or spatial co-occurrence. Here, explicit goal or causal relationships are not necessary. For example, one can talk about the features of an apple, although it may not make sense to ask what the goal of an apple's roundness is or whether the apple's roundness makes it red. We view semantic nets as generally falling within this category although the boundary between semantic nets and object frames is fuzzy along this dimension. For example, a wheel has the property "round" and this feature has a goal or purpose behind it. Normally, we would think of wheel as a single object (hence, represented by a semantic net), yet there is "purpose" behind its features (suggesting an object frame, according to our framework). Hence, organizing principle has limited discriminability between object frames and semantic nets. If knowledge is organized around temporal/spatial co-occurrence and no goal or purpose is present, then it does suggest semantic net. However, if a goal or purpose is present, then it could be either a semantic net or an object frame.

Production rules are also organized around temporal/spatial co-occurrence, in this case, temporal. While goal information may be captured in the antecedent conditions of production rules, they do not influence the overall structure of production rule representations the same way as goals influence the structure of MOPs (goals influence, among other things, the hierarchical structure of MOPs, while production rule chains have the same flat, temporal structure regardless of the goals implicit in them). Similarly, production rules do not have explicit causal information represented (although causal information may be implicitly represented).

Organizing principle is, therefore, a good discriminator of mental models from other structures and between MOPs, production rules and mental models. It may have some diagnosticity with regard to object frames and semantic nets.

The final discriminator we discuss is inferential use. This can be a very powerful discriminator since different knowledge structures are used differently to generate inferences. In classical production systems, inferencing proceeds by matching antecedents of rules to data ("forward chaining"), not by working backward from consequences. To accomplish the latter, a new rule would have to be introduced. MOPs on the other hand support both forward and backwards inferencing in an economical fashion. For example, if one is in a restaurant then, knowing that one has ordered enables the prediction that one will pay later, while observing someone paying allows the inference that they ordered. Hence, unidirectional vs multidirectional inference capabilities is an important discriminator between MOPs and rules. Further, since MOPs are goal-based, these goals can be used in inference, i.e., to explain events. We would expect MOP-based actions to support higher goals--those within a scene supporting the subgoal(s) associated with the scene, while scenes themselves support the main MOP goal (e.g., in the Restaurant MOP, the order scene supports the main goal of eat by conveying one's choice/goal to the person who can satisfy the goal, while reading the menu supports the subgoal of conveying one's choice by informing the actor what the pool of choices are).

Mental models (as we conceptualize them), also have multidirectional capabilities. Causal analysis can be used to generate predictions (forward chaining) or explanations of events (backwards chaining).

Inferencing using object frames and semantic nets is somewhat different than the other three structures. Since objects frames and semantic nets have objects, rather than events as basic units, it makes less sense to think about temporal (forward and backward chaining) inferencing. If anything, spatial inferencing may be more appropriate. We note two types of inferencing that object frames and semantic nets are particular useful for.

First, given one object (in a frame) or an attribute (in a semantic net), one can infer other objects or attributes, respectively, as well. For example, in a tank division, given that we see air defense elements, we might also expect to see artillery (in fact, Leddo et al., 1988, found that intelligence analysts did reason this way).

A second type of inferencing has to do with hierarchical and categorical relationships. For example, if we know a tank has a gun, then we expect to see Abrams tanks having guns as well as well. Further, if we know that a division is a type of combat unit, then we would expect combat units in general to have things like artillery, air defense, etc. Hence, object frames and semantic nets both have multidirectional inferencing, although the types of inferences that are made are qualitatively different than MOPs or mental models. However, objects frames (and perhaps to some extent, semantic nets) are similar to MOPs in that inferences can be made regarding goals or intentions.

Content Knowledge of These Structures. Once knowledge elicitation has discriminated between the different structures, it is important to fill in their content. First, we describe the content (slots) for each of these structures. Here, we rely heavily on the existing literature to specify what the slots are.

MOPs have the basic slots outlined in Schank (1982), and Schank and Abelson's (1977) scripts. Each MOP is a collection of scenes (event sequences) in pursuit of a general set of goals. Each MOP has a header (the name of the MOPs), some background information, and a set of scenes with corresponding actions. The background information includes the list of MOP goals, a set of entry conditions (conditions which trigger the MOP), a list of roles (players in the MOP), a list of props (tools, equipment, etc. used), and a statement of outcome conditions which describe the state of the world (in particular, the status of the goals) after the MOP has been executed. Each scene occurs in a fixed temporal sequence and contains a header and a list of actions. In addition, MOPs also contain information as to how scenes might vary in different situations.

In our framework, production rules have a similar format to that commonly used in the psychology and artificial intelligence literature, with a couple of major modifications. Production rules have antecedent conditions which trigger the rules and consequent actions which are procedures that fire when the antecedent conditions are met. We propose to modify the antecedent conditions slot in two ways. First, the slot is modified to allow multiple possible actions to be fired, e.g., If A, Then B or C. Typically, production rules are deterministic in that a fixed sequence fires when antecedent conditions are met. However, there are numerous conditions where more than one outcome

is possible (e.g., If a coin is flipped, Then it will land heads or it will land tails; If you receive the check, Then pay by cash or pay by credit card). Second, the slot is modified so that actions are fired probabilistically (this is necessary when multiple outcomes,including whether or not to fire the rule at all, are possible).

Semantic nets are used as depicted in the psychology literature, consisting of nodes (concepts) and arcs (relationships between nodes). Examples of arcs include "isa" (one node is an instance of a more general category signified by another node, e.g., Abrams "isa" tank), has-part (one node is part of another, e.g., tank has-part gun), has-property (describes a semantic attribute, e.g., tank has-property fast), etc.

Object frames and mental models have no standard set of slots as far as we know. For mental models, we adopted a set based on the function we view mental models as serving. These slots are context (contains information regarding conditions under which mental models are appropriate), objects (include both animate and inanimate objects), forces (include motivational forces like goals as well as physical forces such as gravity), force/object interactions (the heart of the mental model, which describes how the forces act upon each other and the objects), and outcomes (describes the events or states of the world produced as a result of the force/object interactions).

Object frames contain slots for the objects, how they are configured, what goals (missions) are served, and what more general categories of frames they relate to (e.g., a motorized rifle division is a kind of combat unit, etc.). Our research suggests that people do not seem to have fixed slots for object frames. The slots seem to depend on the context and the purpose the knowledge serves. Hence, our notion of an object frame is less canonical than the other structures.

<u>Methodology.</u> Given that we have discussed the issues relating to the use of CSA, we now discuss how it is implemented during knowledge elicitation. We begin by giving a general overview of our procedure and then discuss the probes we use in more detail. Essentially, we use an interview format in which we embed other knowledge elicitation techniques such as problem solving (typically done small scale with "mini-problems"). The interview portion elicits knowledge at a more general level while techniques such as embedded problem solving illustrate and flesh out specific pieces of knowledge.

In general we employ a "double funnel" methodology for eliciting knowledge using CSA. We use the term "double funnel" as a metaphor describing the open-endedness vs close-endedness of the probes we employ. Hence, the flow of probes is open-ended, close

-ended, close-ended, open-ended. The corresponding steps are general discussion (open-ended), identify structure (close-ended), flesh out structure (close-ended), generalize (open-ended).

We start out at a general level of discussion using open-ended probes such as "How do you do your job?" As we begin to get some content, we look for potential structures that the subject matter expert may be using. This is done based on the discriminators outlined earlier. Once we have established a hypothesis about what type of structure the expert might be using, we begin asking more close-ended questions (thus "funneling" or constraining the subject's responses). At first, these probes are designed to confirm or disconfirm our hypothesized structure by using the discriminators we discussed earlier. Once we feel comfortable that we have identified an appropriate structure, we then ask close-ended questions designed to fill in the slots of the structure. At this point, we often insert a "mini-problem" to have the expert illustrate what he is talking about (e.g., if the expert is talking about how he would prioritize certain targets for firing, we might give him a list of targets and ask him to prioritize them while thinking aloud to articulate his process). Once this is done, we summarize what we have elicited to the expert in order to make sure that we have correctly understood him and try to generalize the specifics we elicited to more general cases (by asking the expert for generalities). We then proceed along the same general lines of discussions we began with. This "double funnel" approach gives the elicitor the flexibility to elicit breadth vs depth knowledge since he can control whether he probes the expert with open-ended or close-ended questions.

Having given this general overview of our methodology, we present detail on the probes we use. In the initial open-ended segment of CSA we ask general questions like, "Tell me how you do your job; what are your objectives; and what information do you use?" Essentially, the specific content of these questions is less important; the main objective is to get the expert talking. From the content of what the expert says the elicitor tries to identify the structure the elicitor is using. This process can be viewed as a decision tree or discrimination net. The general formula for generating probes is to pick a discriminator (such as whether the structure is organized around events or objects) or a slot (such as antecedent conditions of a production rule) and generate questions designed to supply information regarding where the structure falls with respect to the discriminator or the content of the slots. Hence, the specific probes themselves are less important than the knowledge they are designed to elicit. Sample probes are presented to illustrate the principles involved.

The elicitor begins by assessing whether the expert is describing a process or a state of the world. If the elicitor is unclear he can ask the expert. A process suggests either a MOP,

production rule chain or a mental model. A state of the world implies an object frame or semantic net (it could also imply a mental model but only if the expert states how that state of the world is produced, in which case a process would be involved as well).

The next branch of the tree is a further discriminator. For state of the world the next discriminator is whether the expert is referring to a single object or a set of objects. Again, the general formula of asking the expert applies if the elicitor is unsure. A single object implies semantic net while collection of objects implies object frame. At this point the elicitor has identified a hypothesized structure. The next set of probes attempt to elicit content knowledge (fill slots) and validate the structure using the inference-based discriminators discussed earlier. This validation is important to test whether the elicitor has proceeded down the wrong path.

For semantic nets, the next set of probes include: "What are the attributes of this object; why does it have those attributes; give some examples and distinguishing characteristics; what general class does this object belong in; what kinds of contexts does it occur in, how is it used in these contexts?"

For frames, the probes include: "What is the goal or purpose of [the frame]; what are the objects involved (this could then lead to an elicitation regarding individual objects, in which case the semantic net probes would be appropriate); how are the objects related/configured; what context are they found in; what are they used in conjunction with; if you see [object], what do you predict will happen (this tests whether the structure may be event-oriented and is perhaps more suitable for a MOP-based representation); if you see [object], what else would you expect to see; if you see [object], how would you explain it (we would expect reference to the goal)?"

However, if the expert had originally been describing a process and we had hypothesized MOP, production rules or mental model, we would have to discriminate further along these lines. The next probe here would be whether the expert is describing a sequence of procedures or a model of why events happen. If the expert is describing a sequence of procedures, our hypothesis becomes MOP or production rules. If the expert is describing a model of why events happen, we hypothesize mental model.

For mental models, our slot filling probes include: "What contextual elements are present; what objects are involved; what are the causal elements; how do these elements interact; give some examples; how do these parameters vary and how does this affect outcomes; what would you predict might happen given these parameters (here a mini-problem might be appropriate); given this outcome, how did it occur?"

In order to discriminate between MOPs and production rules, we have two initial probes: "Are the procedures based on an overall goal?" and "Do these procedures typically go together and in a fixed sequence?" For MOPs we would expect the answer to these two questions to be yes and for production rules, no. A split may require further probing.

Since production rules and MOPs both describe procedural knowledge some of the probes are identical for slot filling. These will be presented last and shown how they fit in with the other probes. For MOPs the slot-filling probes are: "What goals do you have; with whom do you interact; what tools do you use; what conditions trigger this action; what happens when you're through; describe the procedures you go through; which of these occur in a common setting with similar people (probe is designed to find scene boundaries); what subgoals do you have (for scenes); given [event], what can you predict will happen; given [event], what can you infer is also true or has already happened; why do you do [procedure]; given [event], what caused it?"

For production rules, the slot filling probes are: "What conditions do these procedures occur under (if enough occur in a common setting, it may suggest a MOP is present); describe the procedures (this probe is similar to the MOP probe and is expanded below); how are the procedures related, e.g., does the completion of one procedure trigger another?" At this point, if the expert lists much in the way of relationships, the elicitor may wish to start probing for a MOP and go to the MOP probes that start with "What (sub)goals do you have?"

For describing procedures, there is a set of probes common to MOPs and production rules. These are: "List each step in detail; is the order of the steps fixed?; what information are you acting on; what's guiding what you do (reference to goals?); if the situation were different, what would you do?" While the expert is describing procedures in detail, inserting a mini-problem might be useful to see how the expert actually carries out those procedures.

Once this close-ended probing is completed, the elicitor begins to start generalizing what the expert has stated. Probes for this include: "Suppose we knock out this step, then what; what if we changed the order; what variations are there; how does this fit in with the rest of your job; what would come next?" Included in this process would be a summarization by the elicitor (e.g., "If I understand it, you've described a three step process, the first step being..., etc.") which the expert is asked to validate. Once the expert has resumed his general discussion, the double funnel process repeats itself.

Problem Solving Techniques

There are several variations to the problem solving techn ie. In particular, we identify three dimensions on which to vary the problem solving technique. First, problems can be context- or data-driven (this will be explained below). Second, the problem can be solved by a single expert or multiple experts. Third, the elicitor can use a full-scale problem or a "mini-problem." Actually, this third dimension represents a continuum rather than a dichotomous category.

We begin by describing how these different variations influence the selection of problems and subject matter experts and give guidelines for deciding which variations to use.

Context-Driven Versus Data-Driven Problem Solving. The extent to which a problem is context-driven or data-driven refers to the amount of background or context information that is provided to the expert. In data-driven problem solving, the expert is given (in the extreme case) no background information other than a particular goal (e.g., develop a prediction of enemy intention). The expert must infer from the data he is given the relevant context to his problem. In context-driven problem solving, the expert is given relevant background information (e.g., what unit he is in, whom he is fighting, what terrain, history of the battle) in addition to the goal. Of course, the amount of background provided can be varied according to the elicitor's goals and resources available for problem construction.

The decision to provide or not to provide context entails several tradeoffs. Data-driven problem solving is a much less constrained and structured than context-driven problem solving. For this reason, the expert is more likely to reason divergently and generate creative solutions and inferences. Since the context is not provided to the expert, he must generate inferences regarding the appropriate background information and how they fit in with what he is doing. As a result, this inference process is likely to be more explicit than if the context were already provided and the expert could draw upon it as needed. The drawback with data-driven problem solving is that it is somewhat artificial in that experts rarely operate in a vacuum. As a result, the general problem solving strategy that experts use under this technique is likely not to match how they would approach problem solving under normal problem solving circumstances. However, the data-driven problem solving technique should be rich in terms of eliciting the types of inferences and hypotheses that experts can generate from data.

Providing context to the problem both adds realism to the problem solving process and constrains the number of different inferences and problem solving approaches the expert will

generate. Further, our experience with this technique suggests that there tends to be greater consensus across experts in terms of problem solving approach than with the data-driven problem solving technique. The strength of the context-driven technique is that it enables the expert to utilize his preferred problem solving strategy given the features of the problem. (While the context itself constrains the expert, it is arguable that most real-world problems are constrained and that experts normally operate within such constraints. The goal of the context-driven problem solving technique is to present the expert with a realistic problem and have him solve it in his most natural manner.)

The key differences between context-driven problem solving and data-driven problem solving can be summed up as the context-driven technique constrains the knowledge that the expert will use to solve the problem (thus the elicitor has the flexibility of tailoring problems to elicit specific kinds of problem solving knowledge) and it allows the expert to use his most natural problem solving strategy (given the constraints of the problem). This technique emphasizes eliciting process knowledge rather than content knowledge as the expert will often gloss over details and generate inferences implicitly which go unarticulated.

The data-driven technique focuses on data-driven processes, i.e., the types of inferences and data integration that experts perform during problem solving. Hence, this technique is rich in eliciting content knowledge. Unfortunately, the technique is somewhat artificial and experts often have trouble responding to the probes. While the technique may be useful at eliciting problem solving processes at the micro-level (e.g., how the expert generates inferences), it is poor at eliciting more general problem solving strategies that the experts use in their everyday problem solving.

In choosing which variation of the problem solving technique to use, the elicitor must consider three factors: his goals; the availability of subject matter experts; and the availability of detailed scenarios. With respect to goals, the elicitor must decide whether his main interest is in eliciting process versus content knowledge. The context-driven technique is excellent at eliciting problem solving strategies/processes, but poorer at eliciting the underlying content knowledge. Data-driven problem solving is much richer in terms of eliciting content knowledge and inference generation but is much poorer at eliciting the expert's general problem solving process. Further, data-driven problem solving is more likely to evoke individual differences across experts and elicit creative problem solving strategies, whereas the context-driven technique is more likely to achieve consensus across experts and elicit more standardized problem solving strategies. In the absence of explicit goals, we recommend that

the elicitor begin with context-driven problem solving in order to develop a general framework of how experts in the domain reason and set up problem solving and then conduct data-driven problem solving on the specific aspects of the problem solving knowledge base that remain unfilled by the context-driven problem solving technique.

The second major factor to consider in whether to use context-driven or data-driven problem solving is the type of expert available or that the elicitor wishes to elicit knowledge from. Given that the context-driven technique is more useful for eliciting process knowledge and the data-driven technique is more useful for eliciting content knowledge, it suggests pairing techniques with the type of expert available for elicitation or that the elicitor wants to build a knowledge model of. For example, some experts naturally work at the data level as part of their jobs, e.g., order of battle analysts. We would expect that these experts would be ideal candidates for the data-driven technique. (While such experts may also be useful for the context-driven technique, their most unique contribution is likely to be in the way they process data, generate inferences, test hypotheses, and build a "picture" of the situation they are assessing.)

Other experts work at a higher level of abstraction and may be responsible for integrating the different conclusions of others, e.g., the G-2. Such experts may not focus heavily on specific data, but their merit lies in their ability to generate abstract plans and form general impressions of a problem area. As such, we would expect the expertise of such experts to lie in the processes that they perform. This would make them ideal candidates for the context-driven problem solving technique (which is not to say that they do not have content knowledge that is also useful to elicit).

Given that the elicitor can specify what type of expertise he is most interested in eliciting, he will be able to select appropriate experts and techniques. Fortunately, this is likely to be the case the majority of the time. However, there are times when the elicitors goals may be less well defined or he cannot identify the necessary skills ahead of time. In such cases, the elicitor may wish to follow the guideline offered earlier of selecting the context-driven technique first and then supplementing it with the data-driven technique. If the elicitor selects techniques before he has selected an appropriate expert, then he may wish to select experts based on their appropriateness to the chosen elicitation techniques. For example, if the elicitor plans on utilizing the context-driven technique, then he may wish to select experts whom he believes will have process-related knowledge (e.g., high-level integrators and planners). If the elicitor plans on utilizing the data-driven technique, then he may wish to select experts with more content-oriented knowledge (e.g., data analysts).

The third consideration for whether to use the context-driven or data-driven variant is the availability of resources and/or materials for problem construction. Context-driven problems are more expensive to develop because of the necessary background material and realism. In the extreme case, the data-driven problem solving technique can be performed with limited data and depend heavily on the expert hypothesizing what information he might have and having the expert work from there. While the elicitor can ask the expert to generate material for a context-driven problem, experience suggests that the material the expert generates tends to be sparse and the problem becomes oversimplified. As a result, the expert's knowledge is not truly tapped. It is far easier to ask the expert to generate the next piece of critical knowledge he would need than it is to ask the expert to generate large segments of a scenario.

Full-Scale Versus Mini-Problem. Another decision the elicitor needs to make is the scope (size) of the problem he will use. In essence, this dimension is continuous, although we will discuss this decision in more dichotomous terms, i.e., full-scale versus mini-problem. A full-scale problem is analogous to a complete training scenario where the expert plays a specific role and essentially performs all the steps and functions he would as part of his regular job (or some major aspect of it). Mini-problem focuses the expert on a particular skill or output. For example, a full-scale problem may have the expert play the role of a division G-2 and decide where the enemy main attack may come. The mini-problem may have the expert conduct a force ratio analysis (which might normally be a component of the more general situation assessment problem).

Clearly, there are advantages and disadvantages to using full-scale and mini-problems which the elicitor must consider in choosing which to use. Full-scale problems capture a more global picture of what the expert knows and how he performs his duties. Full-scale problems also enable the elicitor to see how the experts job interacts with other domains, people, etc. Hence, it is easier to get an integrated view of the expert's domain using full-scale problems.

The drawbacks with full-scale problems are that they are expensive to develop and very time-consuming to administer. While these costs are typically offset by the benefits, such problems may be inefficient when the elicitor wishes to probe specific aspects of the expert's knowledge. In such cases, mini-problems may be more appropriate. These allow the elicitor to focus, in depth, on particular aspects of the expert's knowledge. The level of detail elicited per topic is much greater for the mini-problem than it is for full-scale problems. Further, mini-problems are much cheaper to develop and can often be embedded unobtrusively in other knowledge elicitation techniques in order to elicit an illustration of the principles that the expert is presenting.

The drawbacks with mini-problems are their limitations in the breadth of knowledge elicited. In particular, it is difficult to integrate the knowledge elicited using mini-problem into the more general domain knowledge without the help of other knowledge elicitation techniques.

Clearly, the decision on whether to use full-scale or mini-problems will be driven largely by whether the elicitor wishes to probe breadth versus depth of expert knowledge. We recommend a combination of both types of problems and that mini-problems be used largely in embedded format with other techniques. We recommend, in general, that full-scale problems be used first to elicit breadth of domain knowledge and then mini-problems be used to elicit depth of knowledge and flesh out specific areas of the domain.

Single Versus Group Problem Solving. This final dimension refers to how many experts are being used as subjects in a given session. Again, this dimension is continuous. While the number of experts used will influence, to some extent, the materials required, the major determiner of whether to use single or multiple experts will depend largely on the elicitor's goals.

Using single experts allows for the easiest one-on-one interactions. Here, the elicitor can attempt to elicit the knowledge structures of an individual expert. Introduction of additional experts allows for interesting group dynamics. First, experts will often challenge each other, especially when they disagree with each other. This often leads to more "expert" level discussions, particularly when the experts focus on interacting with each other rather than addressing the elicitor (who is typically more naive about the domain, thus inducing the expert to talk "down" to the elicitor). Of course, the elicitor is then tasked with keeping the experts on track and not letting things get out of hand.

Another benefit of multiple experts is having them perform complementary functions, thereby capturing the more group aspects of how experts perform their jobs. This is especially useful in domains where expert knowledge is often captured in group performance as well as individual performance (e.g., the military). Thus group elicitation adds a unique dimension to the elicitation process in these cases.

There are two main drawbacks with group knowledge elicitation. First, it is difficult to capture the knowledge of an individual expert, including conducting the one-on-one probing necessary to flesh out and identify knowledge structures. Of course, there may be cases where the elicitor is interested in institutional, rather than individual, knowledge, in which case the group setting may be more preferable.

A second major disadvantage of the group elicitation process is that it becomes more difficult for the elicitor to maintain control over the elicitation session. In cases where there is a single expert, that expert typically looks to the elicitor for guidance. However, group sessions may generate a dynamics of their own which is independent of the elicitor. If the elicitor can impose a structure on the session, he may be able to regain order. We recommend that group problem solving as a knowledge elicitation technique be reserved for knowledge elicitors with some experience since they may be more used to focusing experts along the lines of specific goals that the elicitors have.

Selecting the Right Combination. We have outlined three different dimensions along which the problem solving knowledge elicitation can be varied: context- versus data-driven; full-scale versus mini-problem; and single versus group problem solving. These dimensions can be combined in any one of eight possible ways. Most of the work in knowledge elicitation has focused on context-driven problems with single experts. However, as we have suggested above, different combinations of these dimensions can be used to build a more complete model of expert knowledge.

We offer some quick rules of thumb for picking values along each of the three dimensions. If the elicitor wishes to emphasize process knowledge, then pick context-driven problem solving. However, if the elicitor wishes to emphasize content knowledge, then pick data-driven problem solving. If the elicitor is interested in a more complete (but less detailed) picture of the expert's skills and tasks, then pick a full-scale problem. If the elicitor is interested in a more detailed view of specific skills, then pick mini-problem. Finally, if the elicitor is interested in the knowledge (structures) of individual experts, then pick single problem solving. If the elicitor is interested in institutional knowledge or group interactions, then pick group problem solving.

If the elicitor's goal is to develop a complete knowledge base of a particular domain, we recommend the following strategy. First, the elicitor should begin with group problem solving using experts of complementary functions. The rationale behind this is that the expert's job typically serves an organizational function and this drives what the expert does. Hence, it is important to understand what the expert does from this perspective. The elicitor should begin with a full-scale, context-driven problems. The elicitor should then switch to single expert problem solving. This will enable the elicitor to map out the general domain and the processes the expert uses. The elicitor can then utilize the data-driven technique, again with full-scale problems and single experts. This will flesh out some of the content knowledge that underlies many of the procedures the expert uses. Next, the elicitor may wish to select specific portions of the knowledge base for enhancing its detail. Here, the use of mini-problems (both context- and data-driven, with single experts) may be useful.

Preparing Problem Solving Materials. Problem solving knowledge elicitation techniques are the most expensive to utilize because of the amount of preparation necessary prior to the elicitation. However, these techniques are also potentially the richest in terms of the knowledge they produce. Essentially, problem solving techniques require that the elicitor provide the expert with a goal and the information necessary to reach that goal. As such, this requires that the elicitor (or his colleagues) have some knowledge of the domain in order to develop realistic problems and appropriate information. Often, the elicitor can supplement his own expertise in problem construction by using domain experts. Another alternative, is to use problems that have been developed for other applications (e.g., training). In the military context, this is often possible given the wide range of scenarios available. If the elicitor chooses this route, he may then need to modify the problem if he wishes to focus on knowledge that is different or more specialized than is intended for the current version of the problem.

We now outline some of the major considerations in problem construction/selection that the elicitor must make depending upon which variation of the problem solving knowledge elicitation technique he will be using.

Context- Versus Data-Driven Problems. The key difference between context-driven and data-driven problem solving is the amount of background information provided to the expert. Clearly, the more information that is provided to the expert, the more constrained he may be in his responses. However, the expert may also be more detailed in his solution and more precise in the knowledge that he uses.

Our experience with context-driven problem solving in military applications suggests that some information is utilized more heavily than others (this will of course, vary by functional specialty and echelon). As a result, we can offer some guidelines to knowledge elicitors using military applications. First, and foremost, military experts working on military problems need to know their current mission. Such information may be presented in terms of the commander's guidance and the operations orders from the echelon above.

Next, subjects rely heavily on friendly and enemy order of battle and current disposition and strengths. Such information is typically provided in G-2 and G-3 estimates. The terrain is typically depicted on the maps and most subjects are content to conduct their own terrain analysis if own is not provided. Subjects in more specialized areas (e.g., order of battler (OB) technicians) may desire Intelligence Preparation of the Battlefield (IPB) products as well. Depending upon the problem, subjects may desire staff estimates from the G-1 and G-4,

information on adjacent units and units of the next higher and lower echelons, and historical information on how the situation/problem evolved to its current state. Our experience with military experts is that they typically prefer visually depicted information, e.g., maps and overlays.

An important part of the problem solving materials is a stream of messages that describe the evolving problem. While problems can be delivered statically with all information presented up front, scenarios realistically evolve over time. As a result, experts will often solve a problem in terms of the directions they see the problem evolving (below, we outline a knowledge elicitation technique called "project ahead" that elicits how experts project different problem evolutionary paths). Hence, the message stream should focus on the key events the expert needs to know about in order to solve the problem. This message stream will typically be organized by time and source (e.g., a report from subordinate brigade that is 12 hours old).

The message stream offers the elicitor a great deal of flexibility in terms of focusing the expert on specific aspects of the problem (as does the rest of the background information) and influencing what reasoning skills the expert will use. For example, incomplete information will force the expert to tap his inferential knowledge. Excessive information will force the expert to generate priorities in terms of what information he will focus on and what information to give the most credence in cases where the information is conflicting. Detailed information may induce more analytical reasoning, while general information may induce more holistic or intuitive reasoning. Here, the elicitor should consider what his goals are in designing the message stream (and more generally, the background information).

In essence, the above description covers the materials necessary to conduct knowledge elicitation using the context-driven knowledge elicitation technique. The data-driven technique eliminates the need for the context information (although the elicitor should have an appropriate context in mind in order to develop coherent data to present to the expert).

In general, the message stream should be rather extensive in the data-driven knowledge elicitation technique. Initially, the data in the message stream should be rather low-level so that the expert does not develop a context too quickly. Recall that the purpose of this technique is to chart the expert's inference process. If the expert infers the context too quickly, then he is likely to generate many implicit inferences which the elicitor may not be able to capture.

In generating the message stream, the elicitor can influence the expert's reasoning processes as well. For example, if the messages relate to a specific topic, then the expert may be able

to quickly put together a small piece of the puzzle (e.g., if the messages relate to the disposition of the enemy's air defense, then the expert may be able to build this piece of the puzzle quickly). However, if the messages are scattered across several parts of the problem, the expert must reason more integratively and try to develop a larger picture of problem.

An advantage of using the data-driven problem solving technique is that the elicitor does not need to develop an exhaustive database of messages. As the expert is being fed the message stream, he will naturally start considering what information he would like to have (in fact, as discussed below, part of the procedure for conducting the data-driven problem solving technique is to ask the expert what additional information he would like). Hence, when the elicitor runs out of messages, he can then ask the expert what information he would most like to have next.

Full-Scale Versus Mini-Problems. Constructing full-scale problems is very different than constructing mini-problems for the expert to work on in knowledge elicitation. Clearly, the latter requires much less effort. As discussed earlier, the elicitor has different goals when using full-scale versus mini-problems. With full-scale problems, the elicitor is seeking breadth of knowledge; with mini-problems, the elicitor seeks depth of a well-defined topic.

When constructing full-scale problems, the elicitor must be sure to include information relevant to the many parts of the problem. Omitting such information will typically result in the expert not considering certain factors and therefore distorting how the expert would realistically solve the problem. For example, many military problems downplay certain aspects such as logistics, air combat power and electronic order of battle. The more the problem leaves out, the more simplified the expert's problem solving becomes and the less useful the resulting knowledge. Unfortunately, full-scale problems are very expensive to generate and require a great deal of expertise on the part of the generator. Our experience suggests that most existing problems tend to focus on particular aspects of the military decision making process, and hence often do not include all of the relevant components. However, we believe that one of the key factors that distinguish experts from novices is the experts' ability to integrate all of the many different facets of the problem into a unified solution. Hence, a person's expertise may truly not be tested unless he is presented with full-scale problems.

A caution must be given with regard to developing and utilizing full-scale problems. Our experience suggests that when given a broad and extensive database, a single expert will often

process the different aspects at a general level. Hence, it is probably reasonable to develop full-scale problems which are not rich in the amount of details the expert needs to process (unless, of course, the expert is given a large amount of time to process such information). In the real world, such richness is handled by distributing the work across several problem solvers. Military decision making is characterized by such a group process. Specific people are then assigned an integrating role, i.e., they take the output of a wide variety of sources and integrate them at a general level (situation assessment and order of battle analysis are examples of functions which operate like this).

We therefore recommend that full-scale problems where the elicitor is interested in the depth of the problem solving process as well as the integration piece be performed in the context of group problem solving. If the elicitor wishes to focus on the integration of multiple sources of information, then using a single subject in an appropriate role would be suitable.

Mini-problems focus on much more limited topics and typically emphasize particular skills. An example of a mini-problem that we have used is computing combat power ratios. Here, the expert is given a specific set of inputs and asked to come up with a specific response. Mini-problems allow the elicitor to chart the expert's problem solving process in greater detail, often to the point of replicability.

As a result, the inputs that the elicitor needs to provide the expert with is much more detailed. The elicitor then needs prior knowledge of what all these inputs need to be and what reasonable values are. Again, having a subject matter expert participate in the materials generation process could prove invaluable. Occasionally, the expert may require a specific piece of information which the elicitor did not provide. In such cases, the expert can be asked to generate a reasonable value and treat it as an input.

Individual Versus Group Decision Making. Much of the knowledge engineering that has been done for expert systems has focused on eliciting knowledge from individual experts. Group "knowledge elicitation" is often done at decision conferences where consensus building is the goal. The difficulty with group decision making is that it is difficult to map out a coherent knowledge representation. A resulting knowledge model may not reflect any individual expert's knowledge (much the way a pooled sample of responses in an experiment may not represent the pattern of any individual subject).

However, the military is relatively unique in that the bulk of the decision making is a group effort (even though the commander has the final say). While the Army publishes extensive doctrine

(which presumably should produce consensus in problem solving thus making group decision making resemble individual decision making), our experience suggests that while there is consensus on the goals that Army problem solvers have, there is widespread variation in the plans used to achieve those goals.

Interestingly, it appears that the appropriate level to gain consensus (i.e., map out group decision making) appears to be at the level of the particular unit. For example, our experience suggests that I Corps at Fort Lewis does target development somewhat differently than III Corps at Fort Hood. Hence, group problem solving may be most useful at trying to capture the problem solving of groups at different units, rather than a particular function (e.g., target development).

As discussed in the section on full-scale versus mini-problems, group problem solving enables the elicitor to explore in greater depth how different experts perform and integrate their respective functions. This would require more extensive problem development. Of course, the drawback is that it is harder to keep track of multiple experts, particularly if they are all asked to think aloud! This problem may be partly solved by utilizing multiple elicitors who are each responsible for recording the responses of an individual expert.

The easiest case in utilizing multiple experts is to have them work together on the same problem a single expert would work on. Here, the elicitor needs no new problem development and it is easier to keep track of the responses of the different experts. However, the elicitor will get more out of the multiple expert format if the different experts can work on separate functions (e.g., CM+D, OB), thus allowing the elicitor to see how the different functions work together.

Procedures. We begin by discussing general procedures for utilizing the problem solving techniques and then discuss variations based on which version of problem solving the elicitor uses. In general, the problem solving techniques are modeled after the standard "think aloud" protocol techniques commonly used in knowledge engineering. This think aloud format typically refers to a context-driven, full-scale, single expert problem solving condition. Hence, the general procedures outlined below apply to this condition. The sections following this description outline how the procedures would change for data-driven, mini-, and group problems.

We find it critical to give the expert a specific goal and have him generate a specific product such as a collection plan, a targeting plan or an estimate of the situation. If the expert is merely asked to demonstrate how he works the problem or is convinced that the elicitor is mostly interested in observing the

process he goes through (which is typically what the elicitor is interested in), the expert will invariably lapse into a description of the steps he takes rather than actually going into the process of solving the problem. If this happens, the elicitation turns into a lecture where procedures are discussed in general terms and no real problem solving is done. The elicitor must therefore emphasize that he is interested in seeing the product. If the session lapses into a discussion or a lecture, the elicitor must remind the expert that he wants the product at the end of the session. Prompts such as, "Let me see how you would actually do what you're describing in the context of this problem.", may be useful.

Once the expert is given his goal, the elicitor asks the expert to think aloud and articulate his thinking process. Such guidance includes asking the expert to describe what information he is using, why he is using it, what conclusions he is drawing, etc. In addition, we typically ask the expert to identify information he would like to have but is not given to him. Here, we can either improvise or tell him it is not available. However, learning what information the expert wants is useful in refining the problem.

Experts will vary on their ability to articulate there thinking; most find it unnatural and will lapse into long periods of silence. This is particularly true when the expert is performing intensive analysis which compete with verbalizing for his attention. The elicitor may find it useful periodically to urge the expert to keep articulating his thought processes. If the expert finds the articulation process unnatural, then periodic urging will realistically do little good. We have found as a useful alternative periodically to stop the expert and have him recapitulate what he was doing and thinking about. Here, the elicitor must be careful not to interfere with the expert's train of thought. Therefore, we recommend that the elicitor stop the expert during "lull" moments, e.g., after the expert appears to have completed a piece of analysis or processed a report.

Periodically stopping the expert tends to work very well in eliciting reasoning processes, even with experts who have trouble articulating their thoughts. Of course, the downside of this approach is that the expert may have to reconstruct his thinking and there he may be incomplete or even inaccurate (cf., Nisbett & Wilson, 1977).

Regardless of whether the expert is thinking out loud sufficiently well or whether the elicitor needs to stop the expert for summarization, we recommend that periodically, the elicitor present the expert with a summary of what he understands the expert to be doing. This helps avoid misunderstandings since the expert has the opportunity to respond to the elicitor's interpretation of the expert's behavior.

We recommend using these recapitulation moments for the elicitor also to ask the expert to clarify any points the elicitor did not understand. The elicitation process loses its validity if the elicitor misrepresents what the expert has said or done. Hence, we recommend that the elicitor seek clarification from the expert even if it means occasionally interrupting the flow of the session (although we recommend telling the expert ahead of time that the elicitor may stop to ask clarifying questions. We also recommend trying to be as least disruptive as possible when this happens.). Another potential risk of asking the expert for clarification is that the expert may begin to infer that the elicitor is naive about the domain and therefore begin to simplify what he is doing so as not to confuse the elicitor. The elicitor must try to avoid this happening. If the elicitor is able to seek clarification after the session (e.g., he has access to other domain experts or can ask the subject), this is desirable. However, if the elicitor deems that understanding the expert's current point is critical to how the elicitor will conduct the session, then the elicitor must seek clarification, in spite of the risks discussed above. We recommend that as part of the clarification process, the elicitor summarize to the expert what he believes the expert has said or done. This helps to reduce misunderstandings.

During the course of the elicitation session, the expert may express the desire to interact with another staff member. Given that the Army context is characterized by group decision making, it is natural for the expert to expect such interaction which is typically for information and analysis. Hopefully, a well-constructed problem will eliminate the need of much of this interaction. However, our experience suggests that different experts interact with other staff differently and it is difficult to foresee what these all will be. Hence, the elicitor needs to prepare for this contingency.

Unless the elicitor is focusing on group problem solving, he does not want the session to develop into having the expert working with another person. However, it is important that the expert receive the necessary input to solve the problem. Hence, the necessary analysis or information should be available. Ideally, the elicitor would be able to provide such information, even if he had to develop it on the spot. However, many elicitors are not subject matter experts themselves. In such cases, we recommend having another subject matter expert serve as co-elicitor. The co-elicitor can provide the subject with the necessary information and analysis. If a co-elicitor is not available or he cannot provide the information, then the elicitor can tell the subject that the information is not available or ask the expert to generate his own assumption (which he then articulates) and work the problem accordingly.

A co-elicitor can serve other useful benefits. The co-elicitor may be useful in administering the problem background (in context-driven problems) and answering any clarifying questions the expert may have. Second, the co-elicitor may add credibility to the session since the expert may feel that the elicitation team will be able to understand what is elicited and not misrepresent it. Finally, a co-elicitor may induce the expert to work at a more "expert" level. If the elicitor is perceived as naive about the domain, the expert may discuss "introductory" aspects of the domain knowledge for fear the elicitor would not understand the more complex aspects.

<u>Data-Driven Problem Solving.</u> Because of its unique format, the data-driven problem solving technique has a specialized set of probes that are presented to the expert. Also, the flow of the technique changes as the expert has acquired information and begun to form a somewhat coherent picture. Initially, the data-driven problem solving techniques focuses on inferencing from specific pieces of data. As the pool of data and inferences grows, the technique shifts to emphasizing integration of the information and developing and testing hypotheses. There is no clear formula for making this shift. The elicitor must judge how well the expert seems to be formulating an overall picture versus still restricting his comments to the individual pieces of data. With experience, the elicitor should learn to make the transition more smoothly.

The data-driven problem solving technique begins with the elicitor presenting the expert with a single piece of information, which could be in the form of a report that the expert might receive in an exercise. The elicitor asks the expert what the report means to him, what he can infer from it. Essentially, the elicitor will continue to press the expert until the expert claims he has exhausted his inferencing. Our experience suggests that the elicitor will get too much, rather than too little, inferencing from the expert.

This first question can be supplemented by having the elicitor ask the expert how additional information might effect these inferences. For example, the expert might cite five potential inferences that could be generated from a given piece of information. The elicitor could ask the expert how additional information would support these inferences or whether they might imply new ones.

Next, the elicitor asks the expert what conclusions or assumptions he has made based on that information. This is important because the elicitor needs to know how the expert is generating the context behind the problem.

After this, the elicitor asks the expert how the current piece of data relates to previous pieces (assuming that others have been

given). Here, the elicitor can be somewhat proactive if he has heard the expert generating related inferences from other information. The elicitor can then ask the expert how the current piece of information relates to related previous pieces of information.

Finally, the elicitor asks the expert what additional pieces of information the expert would like to have and why. This lets the elicitor know what the expert's information priorities are and gives insight into where in the problem the expert is focusing on and how he is attempting to solve the problem.

As the elicitor progresses through the information he has to present to the expert, the elicitor should gradually shift the emphasis of his probes. We would expect the elicitor to focus less on questions such as, "If you had other information, how would this influence the inferences you are making?", because the expert would already have a lot of this other information. Rather, the elicitor would be focusing more on asking the expert to integrate current pieces of information with previous pieces.

When the elicitor exhausts all of the information he has, he can then ask the expert what the one piece of information he would most like to have at that juncture in the problem. The elicitor should caution the expert to state a small-scale piece of information comparable to the ones he has been receiving (e.g., "What is the location of the second echelon's command post?" and not "What is the location of the second echelon?"). Here, the elicitor should be able to generate a response to this request (having a co-elicitor who is a subject matter expert also may help here). If the elicitor cannot provide this information, he can ask the expert to generate what he thinks would be a plausible answer and go with that.

Mini-Problems. Mini-problems are much like full-scale problems in terms of how they are administered. They differ largely in terms of quantity rather than quality. For one thing, there is typically much less background that is needed to be presented to the expert. Because of their compact nature, mini-problems are often relatively easy to construct and can be inserted in other techniques.

Mini-problems demand a greater rigor than full-scale problem solving. Recall that the elicitor is attempting to elicit in depth how the expert solves the problem. Often, the elicitor may be trying to elicit the expert's knowledge in sufficient detail so that the expert's problem solving process could be replicated. As a result, the elicitor may need to play a more active role in the elicitation so that he can better track the expert's process.

Here, it may be useful to have the expert generate intermediate products as he sketches out his solution. The

elicitor should be trying to develop an on-line "formula" of what the expert is doing, which he then tries to verify with the expert. If time permits, the elicitor may try to replicate the expert's process and have the elicitor comment on it.

Group Problem Solving. Perhaps the most challenging of the problem solving knowledge elicitation techniques is group problem solving. It is also potentially the most rewarding since it can capture the group dynamics which are central to many decision making contexts, particularly the Army.

The biggest challenge in administering the group problem solving technique is coordinating the responses of the experts. Our experience suggests that, in group settings, experts alternate between working individually and working together. Often there is no warning as to when the experts will be doing either. The biggest challenge here is to record the problem solving of several experts when they are working separately.

There are a few alternatives possible. First, the experts can be asked to task turns summarizing what they have been doing. This disrupts the spontaneous "think aloud" component of problem solving, but it avoids the confusion of having several people trying to articulate their thoughts simultaneously. Also, experts may feel self-conscious articulating their thinking process in front of their peers.

As second method for handling this is to use multiple elicitors each of whom are assigned to a different expert. This has the advantage of allowing the elicitors to monitor "on-line" thinking (provided the experts are not working so close together that there verbalizations interfere with each other. However, separating the experts too much may interfere with their interactions.) Of course, the drawback here is that more elicitors are required thereby increasing the cost of the elicitation.

There are several ways to utilize more than one expert. Part of this decision will be based on who the multiple experts are. If the experts represent the same discipline, then it may be useful to have them work together on the same task (e.g., develop an order of battle). A variant of this is to have the experts try to challenge each other in order to force them to defend and articulate their thinking. This could be strengthened by assigning one expert the role of devil's advocate or engage the experts in a debate (we have found this latter technique particularly effective in cases where one expert is bent on proselytizing a particular point of view rather than solving the problem).

Another way to use multiple experts is to have them play separate roles. This is particularly useful when the experts have complementary backgrounds and would work together under realistic conditions (e.g., OB tech and collection manager).

There is a potential problem which may arise in the group problem solving format. Often, the experts involved will differ in their rank. When this happens, the higher ranking expert may dominate the discussion and the additional expert(s) may be relegated to rather passive roles. This defeats the purpose of having multiple experts. The problem appears to be less severe in cases where the experts perform separate roles than when they work together in the same role. However, it is undesirable whenever it occurs. If possible, we recommend having multiple experts of comparable rank. Also, the elicitor may be able to counteract the dominating effects of higher rank by explicitly soliciting the input from the lower ranking expert.

Project Ahead Technique

The Project Ahead technique is designed to induce the expert to generate scenarios regarding future or even past events. Not only will the Project Ahead technique help elicit knowledge regarding the types of scenarios that are expected to occur in a given situation, the technique is useful in delving into the causal reasoning behind how those events unfold.

Materials. Materials are optional in the sense that project ahead can be done as an abstract, hypothetical exercise. However, the elicitor still needs to have a particular context (e.g., war unfolding) in order for the expert to project scenarios from. The type and amount of materials the elicitor should provide depends upon the goal of the session. As a rule, the more the elicitor wants to focus on particular types of scenarios (e.g., mechanized warfare in the Fulda Gap), the greater the constraints the elicitor needs to impose. Using scenario information is one practical means of imposing such constraints. Similarly, scenario material helps to focus the expert on greater levels of detail, hence the decision to use scenario material needs to be driven by the level of detail the elicitor is looking for--the rule being, the greater the level of detail, the more detailed the scenario should be. Of course, the scenario material should relate to the topics and level of depth of those topics the elicitor wants to get at. Here, having a subject matter consultant is useful in helping to prepare scenario material for the project ahead technique.

It is recommended that some scenario material be presented to the expert. We have found that providing the expert with a map, overlays of the current situation (both enemy and friendly forces) and a brief history of the activities prior to the situation prove

quite useful in stimulating the expert to generate scenarios. These materials are typically the same used in the problem solving techniques described above.

Procedures. The basic procedure involves having the expert project how a given situation might unfold into the future. The elicitor has several options on how he could guide the expert. These options relate to what the elicitor wants to see projected into the future. For example, in situation development, the elicitor may have the expert project only the enemy course of action. In target development, the elicitor may wish to add what the blue response might be in the way of targeting and how the enemy would react. This could lead to a wargaming-type interaction of blue and red forces. The elicitor may wish to focus on only certain parts of the scenario (e.g., movement of the enemy second echelon, use of artillery, etc.). Essentially, the elicitor is free to constrain the scenario in any way, s/he chooses. These constraints should be driven by the elicitor's goals in terms of what knowledge is needed. Again, the expert consultant can help select a focus of project ahead based on goals of the elicitation.

Project Ahead begins by giving the expert an overview of the scenario s/he will be working from. If the expert has already been working on the scenario/problem, the s/he would already have familiarity. In fact, the point in the problem in which project ahead is introduced would normally serve as the context for which project ahead is to focus.

In order to conduct project ahead, the elicitor needs to perform two tasks: get the expert to initiate scenario generation (which requires that the expert understand what's expected of him) and insure that during the scenario generation, the expert is giving the knowledge that the elicitor is seeking.

We offer a suggested method for introducing the project ahead technique which should be adapted to the specific knowledge requirements of the elicitor (we present several options at different points in the introduction which illustrate how the instructions might be adapted to the elicitor's needs):

> "I would like you to take the current situation as presented to you and try to imagine how that situation might unfold over time."
>
> (Option 1.1) "We are interested in the situation that you view as most likely to unfold from the present situation"
>
> (Option 1.2) "We are not necessarily interested in the most likely scenario, although feel free to

describe the one you see as most likely, but rather a set of plausible scenarios which are representative of would could possibly unfold from a situation such as this."

(Option 1.3) "We are interested in a rather unusual, or even surprising scenario, that might unfold from a situation such as this. Feel free to be creative and develop a scenario that you might not normally expect to see."

(Option 1.4) "You mentioned that you would expect to see a certain event happening (e.g., attack in the south, enemy employ a deception) or were considering a particular hypothesis. Could you illustrate what might take place if that event happened or that hypothesis were true (false)?"

"As you are generating the scenario, please think aloud and describe the events to us. We are particularly interested in your underlying reasoning so please tell us why you think the scenario might unfold the way you describe it and what evidence or indicators that you would see, if in fact, the scenario were developing as you project it."

(Option 2.1) "Please focus on the entire situation, i.e., the red and blue forces and also how the environment might interact with the scenario.

(Option 3.1) "We are interested in how this present situation has evolved. Please present a plausible scenario (here the analogous options 1.1 to 1.4 are possible) as to how these events might have unfolded."

As the expert goes through his scenario development, the elicitor should direct the expert according to issues of interest to him (the elicitor). A set of sample probes are presented in Appendix B. These probes can be used as part of a data sheet, where the expert's answers are recorded during the session.

Conflict Resolution Technique

The Conflict Resolution technique is one of the more interesting knowledge elicitation techniques because it can truly challenge the expert. Here, the elicitor is interested in what data the expert reasons with it, the conditions under which the data are diagnostic vs non-diagnostics, and how the expert handles uncertainty.

The philosophy behind the Conflict Resolution technique is relatively simple. First, the elicitor presents the expert with

information regarding a problem and asks the expert to draw conclusions from it. Next, the elicitor queries the expert regarding the conditions under which the conclusions would be true or not true. Here, the elicitor is trying to ferret out the underlying assumptions that the expert is making in his reasoning. This part of the technique is especially important since much reasoning and problem solving is often based on implicit assumptions that remain unarticulated. In fact, one of the drawbacks with a think-aloud protocol technique is that such assumption may never become articulated during the problem solving flow.

Finally, the elicitor attempts to force the expert to "confront" his assumptions by either challenging the validity of his conclusions given the assumptions or the validity of the assumptions given the conclusion. Interestingly, this manipulation is deceptively simple to introduce, yet devastatingly powerful in its effect on expert reasoning. The elicitor simply tells the expert, "I have an infallible crystal ball that tells me your assumptions are true but your conclusions are false. Why is that?" or "I have an infallible crystal ball that tells me your conclusions are true yet your assumptions are false. Why is that?" These questions are typically introduced at the point where the expert has insisted that he has examined all the relevant assumptions and conditions and is certain in the validity of his conclusion. However, when he is in essence told that he is wrong, invariably the expert comes up with new assumptions and conditions that could impact his conclusion. In fact, in one instance, one of our experts was literally 100% sure that his conclusion was correct. However, when the "crystal ball" manipulation was introduced, he readily generated at least half a dozen plausible reasons why his conclusion could be in error. We believe this Conflict Resolution technique has broader applications beyond knowledge elicitation. We feel this technique is a valuable training tool for a variety of problem solving, planning, and reasoning tasks. Such a technique forces people to examine their reasoning and stimulates creative thinking and multiple approaches to problem solving.

Sample probes for Conflict Resolution and potential data coding sheets are presented in Appendix B.

Integrating the Individual Techniques

Each of the four techniques described above are designed to elicit different, but complementary types of knowledge and reasoning skills that the expert has. In addition to mapping out <u>what</u> the expert knows, the integration of these techniques attempts to map out the <u>constructive</u> processes the expert uses in reasoning with different types of knowledge as well as <u>how</u> the

expert represents that knowledge (to the extent that such "representations" exist and can be approximated using current knowledge representation structures).

The key goal behind our integrated methodology then is to uncover the relationships and interactions between different types of knowledge and reasoning processes. For this reason, we argue that our integrated knowledge elicitation methodology is not a rigid, step-by-step sequence of queries, but rather a set of guidelines for eliciting and building links and relationships between knowledge. Hence, the elicitor is not presented with a decision tree of steps to follow, but rather a set of criteria for assessing the knowledge s/he has elicited so far and a set of guidelines for directing future queries. As a result, the individual elicitation techniques that comprise the integrated methodology are not seen as a set of discrete procedures to be performed sequentially, but rather as a set of knowledge elicitation approaches which are invoked according to the demands of the elicitation session and the evaluations of the elicitor. Hence, an elicitation session using our integrated elicitation methodology will typically be fluid, with the elicitor alternating between techniques as s/he sees fit.

In order to illustrate how our integrated knowledge elicitation methodology is implemented, we return to the basic goals of the elicitation process and our INKS knowledge representation framework (Leddo & Cohen, 1989; Leddo, Cardie, & Abelson, 1987). In knowledge elicitation, we are concerned with two primary objectives: (1) how do experts solve problems, and (2) how can the knowledge of how experts solve problems be captured in a way which is both psychologically valid (as close as we can get) and replicable in applications such as expert systems and training.

In assessing how experts solve problems, we are interested in several things. First, we are interested in the goals and objectives that experts pursue. This also includes the priorities and tradeoffs as well as compatibilities and inconsistencies among goals. Next, we are interested in the planning and problem solving process that experts go through to achieve their goals. This complex body of knowledge is the heart of the elicitation process and encompasses many diverse types of knowledge. The planning process involves, among other things, a causal understanding of the situation that the problem solver is currently in and how it relates to his goals, procedures for generating plans to change the current situation to the desired one, actual "canned" plans which represent past steps that were taken in similar situations (which may be adapted according to the specifics of the current situation). We are also interested in the specific steps or procedures that the expert takes during planning and how the depend on the specifics of the planning

situation. We are interested in the concepts and object knowledge that experts use since these represent the "things" (e.g., weapons, equipment) that expert work with. We are interested in the functions they serve and how they fit in the plans. Finally, we are interested in the causal reasoning and rationales that tie the whole problem solving process together. This includes why the plans lead to the goals, why specific procedures are instrumental to the chosen plans, and why the particular objects or concepts being used serve the functions they do. We feel the causal understanding of the planning process is especially important since this will enable us (or a resulting expert system) to replicate the expert or function in novel situations.

The elicitor, therefore, is guided by this general conceptualization of what he is trying to elicit. The elicitor must then be able to keep the knowledge elicitation session in perspective to get a sense of where s/he stands with respect to this goal (or any goal the elicitor has set up for him/herself). The elicitor then assesses what gaps or general directions s/he needs to proceed in. The elicitor must also understand how his repertoire of elicitation techniques relates to the knowledge s/he is trying to elicit. Finally, the elicitor picks the technique and proceeds with it according to the demands of the situation.

We offer some guidelines to the elicitor. In general, a good way to proceed is to get a general sense of how the expert reasons about the domain. This overview should cover basic things such as the expert's objectives, the kinds of situations s/he deals with, the kinds of people s/he interacts with, the equipment s/he uses, etc. This general orientation will help clarify a lot of what the elicitor will witness during the rest of the elicitation sessions. Generally speaking, we have found that CSA is a good technique to initiate the elicitation process to provide this overview. The elicitor should be attempting to sketch out the different types of knowledge outlined above (e.g., goal, planning) as this will serve as a valuable organizational tool for subsequent elicitation.

The responses typically elicited with CSA tend to be general. Hence, in order to assess how the expert really performs his/her job, we recommend a switch to problem solving, preferably with a full-scale, context-driven problem. The problem solving will probably form the heart of the elicitation session. As the elicitor articulates his problem solving process, the elicitor should continue fleshing out the sketch started during the CSA technique.

It is expected that most problems will involve some sort of information stream which the expert must process, whether it be background information (e.g., order of battle) or a message stream. The expert should be encourage to articulate how he is using the information. At this point, one might feel that it

would be appropriate to switch to the Conflict Resolution technique which deals with information analysis. We feel that it is important to wait until the expert has spontaneously finished his piece of analysis first so as not to disrupt his natural problem solving flow. Our experience suggests that Conflict Resolution works best if we wait until the expert completes a piece of analysis (although he does not necessarily have to complete the whole problem) and then introduce the technique on those pieces of information which the expert states is most important (of course, the elicitor can apply the technique to whatever information s/he chooses). A good guideline for when to introduce the technique is to balance the demands of not interrupting the expert with the desire to elicit the knowledge while the analysis is still fresh in the expert's mind.

During the course of the problem solving, the expert may indicate that he is starting to project future events. This is likely to occur after the expert has become familiar with the problem (i.e., read the background information). Of course, the extent to which the expert projects these events will depend on the type of problem. This is a good time to introduce the Project Ahead technique since the expert is thinking in terms of developing the future events. We recommend using the technique while it happens. However, we recommend getting the expert to systematically project what events he foresees and then delving into the underlying reasoning or asking for specific details (e.g., "What will the artillery be doing as the front line units cross the line of departure?"). Again, these projections might be based on evidence they've seen (e.g., "Well, the lead unit is in the north so they will probably advance in the north along the autobahn."). When this happens, introducing the Conflict Resolution technique may be appropriate.

When the elicitor has completed the elicitation of the problem solving process, it is good to pause and assess where the elicitation stands (in general, regular breaks during elicitation sessions are useful, not only to give the expert a rest, but to give the elicitor a chance to evaluate the progress of the session and plan future lines of query). Here, the elicitor should attempt to get a rough sketch of the knowledge elicited. If the elicitor has overnight, this will be helpful; otherwise the elicitor will have to do this during a break. This assessment will give the elicitor the opportunity to decide what loose ends to tie up and what types of knowledge structures to validate (below we describe how to analyze the elicitation sessions). At this point, we recommend switching to the CSA technique to meet both objectives (fill in loose pieces of knowledge and attempt to validate structures). Keeping in mind that we noted that the CSA technique tends to lead to general commentary, it may be useful to have a set of mini-problems that the expert can work through to illustrate the specific points s/he's making. Of course, this

requires some planning on the part of the elicitor to insure that a good supply of such problems is available (a military subject matter expert may help develop these). Finally, the elicitor analyzes the sessions and plans future ones.

Recording the Session Data

Next, we discuss means of recording the session. It is important to record as much of what transpires throughout the session as possible since it will be later used in analysis. There are multiple ways of doing this, and we recommend generating multiple records (although not every means of recording the session need to be used). The most ideal means of recording the session is to videotape it. This provides both a visual and auditory trace. The visual part is important because it serves as a record of what information the expert used and when as well as the visual portion of the analysis. For example, military problems are very map-oriented. Experts will often work at the map and make reference to map features. Unfortunately, these references are often made by pointing rather than verbalizing (e.g., "The enemy will come from here and attack this point."). Unless, these references are recorded, they may be hard to reconstruct.

The next best method of recording the session is through the use of audiotape. Audiotape captures all of the discussion, but clearly leaves out the visual portion. Here, the elicitor must take great care to make explicit every thing the expert implicitly refers to. This includes the sources of information that the expert is using and the areas on the maps he refers to. When audiotaping a session, we always instruct the expert to be explicit in letting us know what information he is using and where he is pointing on the map. Unfortunately, the expert is not meticulous about this. Hence, the elicitor must often interject for the purposes of the tape. For example, the elicitor may say, "I see you're looking at your order of battle workbook. Could you tell me why?" Or if the expert states, "I think the enemy will attack in this direction," the elicitor might interject, "From northeast to southwest, through avenue of approach A."

A third method we have used is to try to generate an on-line transcription. The rationale for this is that a transcription is required when the knowledge elicitation session is taped (it is too difficult to jump back and forth across the tape and try to remember what was said where and how it all fits together) so that the transcript can begin during the actual session. The advantage is that the transcriber is in the session and can often make use of the visual cues that the elicitor is provided. The transcriber can also be made somewhat familiar with the context, hence there is less opportunity for error. (We have found that most

transcribers are naive about the domain and misinterpret many of the technical phrases and abbreviations used by Army experts, e.g., "early morning radars" instead of "early warning radars." Our experience is that a good typist can generally capture most of what is said in the session (particularly if the participants speak slowly, which is not always the case) and produce a transcript which is comparable to a first draft of a transcript produced by someone not at the session but has the opportunity to replay the tape as many times as needed. Clearly, the on-line transcriber produces a transcript on a one-for-one time basis with how long the session lasts. Post-session transcribers typically require more time to generate a transcription given that they continually replay the tape. One caution in using the on-line transcriber--it is important to clear with the expert beforehand that he will not be disturbed by having a transcriber present during the session. Our experience with an on-line transcriber suggests that experts are not distracted (they easily habituate to the transcriber) and that this method works very well.

A final method of recording the session is for the elicitor to take notes. These can be in the form of longhand/shorthand or using a coding scheme (see Appendix B for some examples). The elicitor should realize that he will only capture a fraction of the material covered and should therefore focus on getting the important points covered and the general flow of the discussion. One of the primary purposes of the notes is to organize the elicitor's own thoughts so that he knows where the session has progressed to and can guide him in future directions. We recommend that the elicitor highlight points that he wants the expert to elaborate on or keep track of topics he wishes to raise later.

In addition to the recording methods driven by the knowledge elicitor and his team, the elicitor can make use of products that the expert's generate in the course of problem solving. These products include notes, overlays, and whatever "deliverables" the expert is asked to generate at the start of the session. Of course, the elicitor must take care that these are properly labeled so that they can be linked back to the place in the problem solving session they were generated.

Analyzing the Data

Generating Transcripts of the Sessions

The first step of data analysis is to generate a transcript of the session to be analyzed. Many knowledge elicitors feel that it is crucial for the transcript to be verbatim, including recording all the "um's" and "ah's" the experts utter while reasoning. We do not feel this level of precision is necessary. We believe it

is more important to capture the gist of what the expert has said than every single word. The verbatim record of the expert's knowledge will be preserved in the session recording in case it is needed later.

We caution the reader that a transcriber unfamiliar with the subject area will not transcribe the session record with 100% accuracy. Hence, it is imperative to have someone familiar with the subject area (and preferably who attended the session) proofread the transcript against the record. Here, the notes taken during the elicitation session may be useful to resolve any ambiguities.

It is important that the transcript include both the elicitor's questions as well as the expert's responses. The transcript should be elaborated so as fill in any undefined referents to maps, reports, etc. This transcript, along with the elicitor's notes, any videotapes, and any products generated by the expert, will serve as the basis for subsequent analysis.

Analysis of the Data

How the elicitor analyzes the session data will depend a lot on his goals. In particular, the application to which the knowledge will be used will play an important role. For example, if the application is an expert system, it will be important to construct a knowledge representation of the data in a form that can be used to generate output (or identify the gaps that need to be filled so that output could be generated). If the application is training, it may be more critical to understand the process the expert uses and less important to understand what knowledge structure gave rise to that process.

Below, we outline a procedure for mapping out the knowledge elicited from the expert. This "map" could then be transformed into a knowledge representation suitable for expert system development (we outline some procedures on how to do this). Finally, we outline some evaluation procedures which can be used to assess how well the knowledge elicitation session went and identify future directions that are needed.

The key to mapping the expert's knowledge is to capture the relationships of the different types of knowledge the expert has. In general, we expect experts to have knowledge of their objectives, general planning knowledge for strategies for achieving these objectives, specific procedures to carry out these plans, content knowledge of the situations these plans apply to, and causal knowledge of how these plans and actions achieve the objectives in the specific situations. We argue that this conceptualization allows the knowledge elicitor to set up a general framework for data analysis and that he can map the

expert's responses on to this general template. If the elicitor requires a more tight representation, he can translate this general template into specific knowledge structures.

We recommend mapping out the expert's knowledge from the top down. Essentially, the elicitor should begin by trying to map out the expert's general problem solving strategy which will serve as a framework for the rest of the expert's knowledge. Conceptually, the elicitor is trying to map out the broad steps that the expert followed and then fill in the more detailed procedures and content knowledge the expert used. It is important, to the extent this was elicited, to include the rationale behind the expert's strategy, i.e., the reasoning behind why the expert chose the steps he did.

Mapping out the expert's knowledge is a difficult process since the expert's knowledge will not fall into one neat package. Often, the expert will call upon different bodies of knowledge to support his problem solving. Hence, the resulting map of the expert's knowledge is likely to resemble a diverse set of knowledge structures or packets which are integrated by the problem solving strategy. Because there is a certain amount of skill and specialized knowledge required to partition out knowledge structures and infer appropriate linkages between structures, we recommend that the elicitor have some background in cognitive science. (We recommend that the elicitor perform the data analysis since he was at the session and understands the general context in which the expert's responses were given. Also, there are likely to be certain cues that the elicitor picked up during the session, not captured in the transcript, which help to clarify the expert's meaning.)

The first step in the analysis is to map out the expert's goal structure. This is a set of goal relationships that drive how the expert solved the problem. Different characteristics of goals include what they are, who has them, their relative importance, time dependencies, and goal relationships. Goal relationships are subtle, but very important. Potential relationships include: instrumental (one goal helps to satisfy another); adversarial or mutually exclusive (two goals are in opposition to each other, e.g., both sides want to win the war); subgoal (one goal is part of what is necessary to achieve a more major goal); consistent (achieving one goal is compatible with achieving another).

Many goals will be straightforward to identify and classify. Often information regarding them will occur in direct response to a question by the elicitor. Other goals can be inferred from the expert's comments, e.g., "We need to accomplish this...; we have got to be able to do...; my number one priority is..."

Goal relationships are harder to infer. Hopefully, the elicitor has asked appropriate questions to flesh these

relationships out. If not, the elicitor can follow some guidelines to infer goal relationships. First, the elicitor can assume that the goals of friendly and enemy forces are adversarial. The expert may often generate phrases like, "The enemy wants to do this, which will prevent me from accomplishing that," and vice versa.

Subgoals refer to goals that are components to a more major goal (i.e., the major goal is broken down into intermediate steps). For example, situation assessment may have as subgoals, terrain analysis, develop enemy OB, etc. The expert might cue these with utterances such as, "I need to do this in order to do that; this is part of what I need to do," etc.

Instrumental goals are contrasted with subgoals in that the former can be viewed as setting up plans for achieving the major or subgoals, while subgoals refer to components of the major goal. For example, while developing the enemy OB might be a subgoal of situation assessment, developing a collection plan might be instrumental in developing the OB. Hence, instrumental goals relate to how the expert will achieve his goals, whereas subgoals relates to what the components of the major goals are. Hence, the elicitor infers instrumental goals from the plans that the expert uses to satisfy major or subgoals.

Consistent goals refer to goals which do not oppose each other and are often complementary. For example, an expert may often cite goals of predicting enemy composition and disposition in generating enemy OB. These goals are consistent with each other, but are not really instrumental to each other (although sometimes composition and disposition correlate, e.g., tanks may be held in reserve and mechanized infantry placed forward), nor is one a subgoal of the other. In general, it will take some knowledge of the domain to assess whether goals are consistent with each other or whether they are in some other relationship. Unfortunately, the fact that a single person or organization holds two goals does not necessary distinguish whether two goals are consistent versus adversarial. For example, the goal of successful offensive engagements and the goal of minimizing attrition may trade off (i.e., one must expend lives to accomplish a successful attack).

Once the elicitor has mapped out the expert's goals, he should focus on mapping out the expert's general strategy. Essentially, this involves mapping out the major steps that the expert went through in order to solve the problem. Hence, the elicitor should look through the transcript of the problem solving session and pull out the major steps the expert went through and ignore the details (which will be filled in later). The elicitor is interested in mapping out the processes the expert went through (e.g., compute force ratios, analyze terrain), the order in which they occurred, and any interactions (for example, the expert may

analyze the terrain in a particular sector, look at how the forces are arrayed along the terrain, then repeat the process for another sector).

Next, the elicitor should map out the procedures the expert used to carry out each step in the process. Of particular interest are the content of the procedures (what the expert did), the order in which they occurred (including whether there appeared to be flexibility in the order--as indicated by a response to a direct query or if the expert varies the sequence over iterations of the process), the triggering conditions, the actors (people involved in the procedures), the props (tools, equipment, etc.), any projected outcomes, and any explicit links to the goals.

Next, the elicitor should map out the content knowledge involved in carrying out the procedures. Content knowledge refers to concepts or objects that play a role in the process. Content knowledge involve concrete concepts such as "tanks" or more abstract concepts such as "momentum." Similarly, content knowledge can refer to single or collections of objects. For example, a tank is a single object, but a tank division refers to a collection of objects. In analyzing content knowledge, the elicitor is interested in two things: the qualities or attributes of the concepts and the relationships of the concepts to other concepts and the processes that use them.

The attributes of the concepts will generally be derived from descriptions of them, e.g., tanks have tracks or wheels, they are armored, etc. There is really no predefined set of attributes to look for and, in general, an object's relevant attributes will vary according to the domain they are used in and the way they are used.

Relationships of objects/concepts is a more difficult problem, and here again, there is really no predefined exhaustive set of relationships. There are two broad categories of relationships worth considering: categorical and functional. Categorical relationships refer to semantic categories. For example, an Abrams would be an instance (example) of a tank, which in turn is an instance of an armored combat vehicle.

Functional relationships are more subtle and in many ways, more important. Functional relationships include things like spatial and temporal configurations (i.e., where and when things happen in relationship to each other), and their connections to goals and plans (e.g., in the offense, the configuration of combat support units is brought forward which supports the goal of forward advance of forces).

Finally, causal relationships provide a rationale for why plans and problem solving strategies support goals. Often, these causal connections link content knowledge as well as actions to

goals. For example, setting up defensive positions on heavily wooded hills is often an effective plan for defending forces. Hills provide good line of sight to enemy forces and woods provide good cover and concealment for friendly forces. The rationale for these defensive positions not only relate to the benefits of line of sight and cover and concealment, but also to the attributes of woods and hills for providing these benefits.

In mapping out these causal relationships, the elicitor should be sure that the explanation provided by the expert relates to a specific goal or subgoal. Next, the causal relationship needs to take into account the situational and/or procedural factors involved (i.e., what the expert is doing and/or using to accomplish the (sub)goal). Finally, the causal relationship needs to specify how the situation and procedures leads to the attainment of the goal. It is desirable, if possible, to express the relationship in general terms (even as a formula) so that the causal relationship can be used to explain other actions or even make predictions regarding goal attainment given a similar set of circumstances.

Validating the Knowledge

It is important to validate the elicited knowledge once it has been compiled. Validation is a difficult problem in a domain such as military decision making or analysis in that few objective standards or right answers exist. We attribute the lack of good, objective criteria to several factors: military science is a rapidly evolving domain--new weapon technologies rapidly shape our conception of how war is to be fought; the types of war that the Army is training for have not been fought, hence there is little feedback as to what will work and what will not work. As a result, we argue that most validation will have to been driven by considerations of consensus and logical consistency.

We outline two broad dimensions of validation of elicited knowledge: completeness and validity. In addition, there are efficacy measures that can be applied to the knowledge elicitation techniques themselves. We discuss these in turn.

Completeness. Simply stated, completeness refers to whether the elicitor captured all of the expert's knowledge. Realistically, it is not practical to expect or even necessary that the elicitor capture all of the expert's knowledge. It is more useful to evaluate what the elicitation did capture and where the major gaps are. Of course, if the elicitor had a predefined objective, e.g., build an expert system, then this objective could be used as a standard for evaluating whether sufficient knowledge is elicited.

We offer two measures of validity: expert review and replicability. Expert review refers to the extent to which expert reviewers view the elicited knowledge as complete. Clearly this

is subjective to some extent. However, by using multiple judges and computing interrater reliability, the elicitor can gain some consensus regarding the completeness of the knowledge. The elicitor can implement this evaluation by giving experts a copy of the represented knowledge and ask the experts to critique the knowledge, in particular, assessing whether the relevant goals and procedures are included, where detail is missing, and what major pieces of knowledge are missing. We feel this qualitative evaluation is more useful than a quantitative one (e.g., have the experts rate how complete the knowledge is) for two reascns: first, the elicitor is likely to benefit from feedback which explicitly tells him what knowledge he needs to elicit further; second, judges are likely to differ in how they use rating scales, so while the elicitor may see a correlation in rating scores across subjects, he may have a less good sense of how complete the knowledge really is (e.g., if one expert tends to give high ratings and the other gives low ratings).

Another source of "expert" review might be published doctrine. Here, the elicitor would compare the knowledge that he has elicited from the expert to the procedures that doctrine says the expert should follow. This may be useful in identifying major procedures that the expert leaves out. Unfortunately, this method may be unreliable, particularly if the expert uses another set of procedures which are equally valid or if the expert deliberately omits a step because it is not necessary in the particular situation.

The second major way to assess completeness is replicability. Replicability refers to whether a naive problem solver can solve the same problem as the expert using the summary of the expert's knowledge. Clearly, this is a more rigorous test of the elicited knowledge and one that is necessary if the goal is to replicate the expert as is the case in expert systems.

Here, the expert's knowledge summary is presented to a novice. The novice is given the same problem that the expert's knowledge summary is based on. The novice is asked to solve the problem. While, it is unlikely that the novice will be able to solve the problem and come up with the same answer as the expert without some help, this evaluation technique is useful at identifying gaps in the knowledge summary (as evidenced by the help the novice needs). Of particular value to this technique is the implicit knowledge and assumptions the expert makes, that the novice will not have. Hence, when trying to replicate the expert's solution, the novice will likely expose these assumptions by reaching impasses.

Validity. The other major criterion for evaluating the elicited knowledge is the validity of the knowledge. This refers to whether the expert's knowledge is "right." As discussed above,

this is a difficult issue in domains such as intelligence analysis where it may be unclear as to what the "right" answer is. Again, we propose two criteria for assessing validity: consensus and logical consistency.

Consensus refers to whether it is agreed among other experts that the knowledge elicited is "right." "Other experts" can be individual people, some generally accepted (school) solution to a problem, or a prescribed (doctrinal) way of approaching problems.

One method of validating the expert's elicited knowledge is to present the summary of his problem solving process, along with the solution he came up with to a panel of experts (again, it is desirable to obtain a measure of interrater reliability) and have them rate the quality of the knowledge. Again, qualitative assessments of what is good and bad about the expert's problem solving process and/or solution is likely to be more valuable to the elicitor than a quantitative rating (especially if the elicitor has the goal of using the knowledge base for a particular application and needs to decide what parts of the knowledge base to keep or discard). A variation of this technique is to compare the knowledge elicited from several experts (subjects) in order to gain a consensus.

Occasionally, although not always, there will be an approved solution to the problem presented to the expert, including "correct" inferences and conclusions that the expert should be making along the way. In such cases, the expert's protocol could be compared to this school solution for consistency and to identify discrepancies. The elicitor should be cautioned to look out for cases where other solutions may be acceptable. (Comparing an expert's solution to a school solution presupposes that the school solution is the only "correct" solution. The elicitor needs to verify this before using this evaluation method.)

The third consensus criterion is to compare the expert's problem solving process to published doctrine. Here, the elicitor looks for where the expert conforms to doctrine and where he deviates.

A word of caution should be noted regarding "consensual" methods of knowledge validation. Consensus implies that the "majority" is "right." If one believes that skills are normally distributed, then the best expert will actually differ from the majority (unfortunately, so will the worst expert). Therefore, a potential danger of consensus methods is that the "best" expertise may be missed, unless the majority of experts can recognize superior expertise when they see it. Unfortunately, given that expertise is rarely tested in wars, it may be difficult for current Army experts to recognize the "best" wartime expertise in peacetime contexts.

The second category of validity evaluation is logical consistency. Here, the expert's knowledge is subjected to tests of logical rigor. Such tests include whether the expert contradicts himself and whether the expert's conclusions logically follow from his premises. While such tests may seem straightforward, it is important to realize that sometimes different pieces of knowledge may appear logically inconsistent, when in fact additional unstated premises may resolve the conflict. We have found it useful to present such inconsistencies (both within and between experts) to the experts themselves and ask for resolutions. We have found that sometimes these inconsistencies are genuine, but most often they are resolved through the addition of premises or applicable conditions.

Knowledge Elicitation Technique Evaluation. In addition to the elicited knowledge, the knowledge elicitation technique itself can be evaluated along different dimensions. The broad categories of dimensions that can be used include: expert acceptance, elicitor skill required, efficacy, and robustness.

Expert acceptance refers to the expert's perceptions, feelings, etc. on how the elicitation went. Here, after the elicitation session, we ask for the expert's view on the session. Questions we ask include: "Were the questions we asked you and the tasks we asked you to perform compatible with the way you naturally think about the problem; did you find the problem challenging; do you feel we covered the relevant topics, and if not, what did we leave out; how much do you feel you told us was derived from book learning vs practical experience; what suggestions do you have for improving our elicitation procedure?" We find that the integrated knowledge elicitation methodology (with the exception of the data-driven problem solving variations) is quite natural to the experts and they frequently feel positive toward the elicitation session.

Elicitor skill refers to what training and background the elicitor needs to have to perform the technique. In particular, the two main relevant skills are domain expertise and cognitive science expertise. As discussed above, domain expertise is more relevant when eliciting content knowledge and cognitive science skill is more relevant when eliciting the structure of the expert's knowledge as in the CSA technique. The problem solving techniques elicit more content than structure, hence, little cognitive science knowledge is required and more domain knowledge is useful. Fortunately, the burden of the elicitation is placed on the expert (he is encouraged to solve the problem naturally and think aloud). Therefore, the elicitor can get by with little domain expertise as well.

Efficacy refers to how efficient the knowledge elicitation technique is. This is hard to quantify objectively since a number of dimensions may be relevant and the elicitor has to decide what

is important. One dimension of efficacy can be viewed as knowledge elicited per unit time or how long it takes to perform the technique weighed against what knowledge is elicited. While it is easy to quantify time, it is more difficult to quantify knowledge, especially since there are different types of knowledge (e.g., plans, procedures, etc.). In general, there are more low-level pieces of knowledge than high-level, thus making a comparison difficult. For example, is it more efficacious to elicit two plans or five production rules?

Given that different knowledge elicitation techniques appear to be better at eliciting different types of knowledge, efficacy comparisons may be better performed across techniques for the same type of knowledge. For example, problem solving may be most efficacious at eliciting planning knowledge while Conflict Resolution may be more efficacious at eliciting inferential knowledge.

A second dimension to efficacy relates to the quality or relevance of the knowledge elicited. For example, it may not be efficacious to elicit knowledge which is duplicative of knowledge that can be readily found in a book or elicited from novices. Here, the elicitor must evaluate what his goals are with respect to the knowledge he wants to elicit and evaluate the efficacy of the knowledge elicitation technique with respect to eliciting the desired knowledge.

The final criterion for evaluating the knowledge elicitation technique is robustness. By robustness, we mean how well the technique works across different experts and different domains. One can make this assessment by looking at other criteria presented in this section and how well they are met across subjects and domains. This evaluation gives an indication of how sensitive or limited the technique is.

For example, we discussed earlier the importance of the expert thinking aloud while he solves the problem. Hence, problem solving techniques may be less robust across experts since they require a verbal ability that not all experts have. Similarly, problem solving techniques may be more appropriate in domains where problem solving procedures are common (as opposed to domains where people play "hunches," e.g., dating). Since military expertise is heavily problem solving oriented, we expect (and have found in the four domains tested in the present work) that problem solving elicitation techniques will be robust across domains. Project Ahead should also be robust as the heart of military operations is planning for the future. While Conflict Resolution may be applicable across domains, it is probably most relevant for inference specialists such as members of the G-2 staff. Finally, CSA should be appropriate for all domains.

APPENDIX B

SAMPLE QUERY SHEETS FOR CSA, PROJECT AHEAD,
AND CONFLICT RESOLUTION TECHNIQUES

MOP-BASED EVENTS

Cross-Checking Probes

EVENT:

Sequencing

Describe the sequence of steps that take place during this event?

Is this sequence fixed? Are there any alternatives?

Triggering Conditions/Causes

When does this event occur? (What conditions?)

What has to come before it? (What causes the event?)

Is this prerequisite necessary?

What could prevent the event from happening?

How would that prevent the event from happening?

Indicators/Inference

What are the indicators of the event? How do you recognize them?

What would be indicators that the event would not occur? How do you recognize them?

Outcomes

How does this event change the situation you're in?

What would come next?

PROJECT AHEAD

Probes

Describe the events that you see unfolding.

What is the enemy force doing?

What is the enemy trying to accomplish? (Why is he doing this?)

Step ____ (time occurs):

Component______________________

Why is this happening? (What is the enemy trying to accomplish?)

Location of activity:

Timing (time frame) of activity:

What indicators do you see?

PROJECT AHEAD

Probes

Describe the events that you see unfolding.

What is the friendly force doing?

What is the friendly force trying to accomplish? (Why is he doing this?)

Step ____ (time occurs):

Component______________________

Why is this happening? (What is the friendly force trying to accomplish?)

Location of activity:

Timing (time frame) of activity:

CONFLICT RESOLUTION

Probes

Hypothesis, Conclusion, or Event the expert has listed:

Indicator or piece of evidence number ________.

Conditions under which evidence implies event?

What makes evidence imply conclusion?

What would have to happen for you to see this indicator?

How likely is it that you would see this indicator?

What could prevent you from seeing this indicator?

What other indicators would strengthen conclusion?

What other indicators would weaken the conclusion?

Would ___________________ indicator strengthen, weaken, or be irrelevant? Why?

When does this indicator not imply the conclusion?

What else could this indicator mean? (what other conclusions?)

I have a crystal ball which says the indicator is true and all the conditions are met but your conclusion is false. Why?

I have a crystal ball which says your conclusion is true but the indicator is totally irrelevant. Why?

APPENDIX C

DATA REDUCTION AND CODING SHEETS FOR SUBJECT'S RESPONSES DURING KNOWLEDGE ELICITATION USING THE INTEGRATED METHODOLOGY

The Goal Tree

Whenever a goal is articulated, the following should be coded (to the extent the information is available):

(1) Goal.

(2) Owner (whose goal it is--friendly, enemy, G-3, etc.).

(3) Category (e.g., tactical, operational, strategic...).

(4) Timeframe (e.g., short-term, long-term, number of years).

(5) Importance.

(6) Link to:

Other Goals:

+ interactions (esp note more/less important than)
+ instrumental
+ subgoal/supergoal
+ plans
+ mental models
+ concepts

(7) Annotation (free-form).

Plans

(1) Actions (procedures).

Temporal sequence: fixed/variable/time dependencies

(2) Triggering conditions.

(3) Actors (Who is involved?).

(4) Props (What equipment, tools, resources are used?).

(5) Outcomes.

(6) Contingencies.

(7) Flexibility in sequence/action.

(8) Context.

(9) Link to:

Goals
Causal Models
Concepts

(10) Annotation (free-form).

Causal Models

(1) Objects (people, units, entities, etc.).

(2) Forces (what influences events, e.g., motivation, combat power).

(3) Force/object interactions (how forces interact with objects to produce events).

(4) Outcomes (events produced as a result of these interactions).

(5) Context.

(6) Link to:

Goals
Plans
Concepts

(7) Annotation (free-form).

Concept Knowledge

(1) Objects.

(2) Properties (attributes).

(3) Components (parts).

(4) Instances (examples).

(5) Category membership (category that present concept is an example of).

(6) Other relationships to other concepts (spatial, temporal, etc.).

(7) Links to:

Goals
Plans
Causal Models

(8) Annotation (free-form).